STRUCTURAL TRANSFORMATION IN SUB-SAHARAN AFRICA: AGRICULTURAL PRODUCTIVITY, LABOR SUPPLY, AND OCCUPATIONAL CHOICE

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ellen B. McCullough

January 2017

STRUCTURAL TRANSFORMATION IN SUB-SAHARAN AFRICA: AGRICULTURAL PRODUCTIVITY, LABOR SUPPLY, AND OCCUPATIONAL CHOICE

Ellen B. McCullough, Ph.D.
Cornell University 2017

In the first analytical chapter of this dissertation, I draw on a new set of nationally representative, internationally comparable household surveys, in order to provide an overview of key features of structural transformation — labor allocation and labor productivity — in four African economies. New, micro-based measures of sector labor allocation and cross-sector productivity differentials describe the incentives households face when allocating their labor. These measures are similar to national accounts-based measures that are typically used to characterize structural change. However, because agricultural workers supply far fewer hours of labor per year than do workers in other sectors in all of the countries analyzed, productivity gaps shrink by half, on average, when expressed on a per-hour basis. Underlying the productivity gaps that are prominently reflected in national accounts data are large employment gaps, which call into question the productivity gains that laborers can achieve through structural transformation. Furthermore, agriculture's continued relevance to structural change in Sub-Saharan Africa is highlighted by the strong linkages observed between rural non-farm activities and primary agricultural production.

The process of economic development is characterized by rising output per agricultural worker and the exit of labor from agriculture to other sectors, which together result in rising incomes and falling incidence of poverty. In my second

analytical chapter, I explore the relationship between labor productivity and the occupational choice that underlies the structural transformation process. I model households' decisions to participate in different activities – farming, wage employment, and self employment – through operation of a household non-farm enterprise. I estimate a structural, polytomous model of occupational choice using nationally representative household survey datasets from Tanzania, matched geospatially to several other relevant datasets. Then, I simulate the response of occupational choice to stylized productivity shocks to farming, wage employment, and self employment. I find that participation in farming is not responsive to productivity shocks of any sort. This is most likely because farming participation rates are already quite high. Wage and self employment participation do respond to wage and self employment productivity shocks, respectively. These results highlight the importance of investing in improved smallholder farmer productivity, especially along the intensive margins of farming participation and especially in places with low population density and poor market access, where farming productivity gains are the only ones to impact households.

Investing in productivity-enhancing inputs is complicated by variability in rainfall, temperature, infrastructure, soils, and market access, which condition the economic returns to input use over space and time. Newly available, spatially explicit data in Sub-Saharan Africa allow decision makers to better understand how agricultural production and prices change with this variation in climate and growing conditions. In my third analytical paper, I, along with coauthors, develop an innovative, *ex ante*, spatially explicit profitability assessment tool in order to inform large scale operational decisions in the presence of risk and uncertainty. This tool allows decision makers to visualize the probability of achieving profitability objectives when climate conditions and prices are unknown. We develop this decision tool in Ethiopia, a country characterized by its high levels of spatial heterogeneity, rainfall risk, and price risk, as well as its strong commitment to investment in agricultural growth and transformation. We use a large scale experimental dataset to cleanly estimate the production response to fertilizer application conditional on climate and soil conditions. Using these model parameters, we simulate the profitability of fertilizer use conditional on market conditions at the time of fertilizer purchase. We explore the implications for decision makers who are designing and targeting soil health interventions. Though this decision tool is developed for nitrogen management on maize in Ethiopia, the novel approach can be expanded to other crops, nutrients, and management practices.

BIOGRAPHICAL SKETCH

Ellen McCullough joined the doctoral program in the Charles H. Dyson School of Applied Economics and Management in the fall of 2011. Her research focuses on the productivity of smallholder-based agriculture in Sub-Saharan Africa. Prior to pursuing her Ph.D., she worked as an Associate Program Officer in the Agricultural Development program at the Bill & Melinda Gates Foundation. There, she developed a portfolio of grants focused primarily on generating better household level survey data to guide policies and investments. She also worked to enhance the nutrition impacts of the foundation's agriculture-focused investments. Before joining the foundation, Ms. McCullough worked in the Agricultural Development Economics division of the United Nations Food and Agriculture Organization in Rome, Italy. She led a cross-country research project on organizational changes in food systems and the impacts on smallholder farmers.

Ms. McCullough graduated from Stanford University with a Bachelor of Science in Earth Systems (2003), and a Master of Science in Earth Systems (2004). Immediately following graduation, she worked as a research fellow in the Freeman Spogli Institute for International Studies.

This document is dedicated to those who earn their livings through farming.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge my special committee chair, Chris Barrett, for his intellectual leadership, guidance, and support. I'd also like to thank David Just and Victoria Prowse for their very useful insights as members of my special committee. I received financial support through the Food Systems and Poverty Reduction Integrated Graduate Education and Research Traineeship (IGERT) of the National Science Foundation and the African Development Bank. I chose to attend Cornell because of its strong reputation in international development economics, and I was not disappointed with this decision. I benefitted greatly from interactions with the community of researchers considering international development topics, especially related to agriculture. This includes, but is not limited to, the Chris Barrett research group, broadly defined. I'm grateful to many others in the Cornell community, including Prabhu Pingali, Arnab Basu, Josh Woodard, David Lee, Aija Lieponen, Miguel Gomez, Per Pinstrup-Andersen, Bob Herdt, Rebecca Nelson, Erik Thorbecke, and Clark West.

Above all, I am grateful to my family for their unwavering support, including my parents, Effie & Tom McCullough, my brother Edward, my sister Margaret, her husband Richard, and their son George.

I acknowledge how well I have been served by some of the world's top educational institutions. I am profoundly grateful for the opportunities this education has afforded me, and I pledge to use these opportunities to serve those who have not been so fortunate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Overview

Structural change underpins the economic development process. Structural change refers both to the processes of reallocating labor from low- to high- productivity sectors and of new technology adoption and to the economic growth resulting from those processes. When there is a large productivity gap between sectors, the gains from labor allocation can be large. Similarly, when there is a large productivity gap between incumbent and new technologies, the gains from adoption can be large. Though land productivity growth typically precedes labor productivity growth within the agricultural sector, the process of agricultural development is thought to begin with growth in output per agricultural worker (Timmer, 1988). Agricultural labor productivity typically starts lower than non-agricultural productivity but rises faster due to technological change. Structural transformation is characterized by falling agricultural labor shares as labor then shifts from agriculture to more productive sectors. It is premised on agricultural productivity growth exceeding non-agricultural productivity growth.

A “developed” Africa, almost by definition, will be characterized by a more productive agricultural sector and fewer workers in agriculture. Though the end point is fairly clear, it is not clear how most African countries will proceed on this path towards development. Where and how will productivity growth take place, and who will exit farming and when? These structural change patterns have very important implications for how the gains from structural change are distributed among workers and the owners of capital in the economy. They are also very central for prioritizing interventions, such as technology to improve smallholder agricultural productivity versus education of future wage laborers, in order to maximize the impacts on poverty reduction. Understanding how these changes are playing out in Sub-Saharan Africa, a region on the precipice of major growth and transformation, is essential for managing that transformation.

This dissertation focuses on large scale structural change processes in African economies, in order to develop a better understanding of how poor, largely agrarian countries transition towards more developed and diversified economies.

1.2 Structural change

Productivity growth in agriculture can raise farmers incomes, push up wages outside of agriculture, and pull people out of agriculture. McMillan and Harttgen (2014) attribute declining agricultural employment shares to agricultural productivity growth. In order for the industry and service sectors to absorb new workers from agriculture, their overall productivity must grow faster than their labor productivity. Slow growth in industry and service sectors can therefore suppress agricultural productivity growth (Timmer, 2009). Agricultural growth can also lead to decreases in real food prices. Because consumer preferences between food and non-food goods are non-homothetic, demand for industrial goods and services increases as income grows among smallholder farmers (Michaels, Rauch, and Redding, 2012).

Agricultures posited role in promoting agricultural growth includes creating profits for other sectors and serving as a reservoir for surplus labor (Ranis, 2004). Arthur Lewis famously hypothesized that labor could be shifted out of agriculture without loss to agricultural output (Lewis, 1954b). In a closed, two sector economy, owners of capital in the industrial sector could accumulate additional capital by absorbing workers from agriculture without bidding up wages and without loss to output in the subsistence farming sector. Lewis later extended this analysis to open economies, where he argued that increased productivity of labor for agricultural food crops could boost terms of trade for poor countries with low labor productivity that relied on agricultural exports (Lewis, 1969).

Recently, the structural change literature has seen a resurgence of papers focused on measuring cross-sector labor productivity differentials and their implications for growth dynamics, equilibrium wage rates, and labor allocation patterns (Christiaensen, Demery, and Kuhl, 2011; McMillan and Harttgen, 2014; Gollin, Lagakos, and Waugh, 2014b,a; Rodrik, 2014a). The existence of cross-sector productivity gaps implies a misallocation of production factors in the economy. Explanations for these gaps have emphasized worker self-selection, measurement of labor inputs and their quality, and policy distortions (Ranis, 2004; Gollin, Lagakos, and Waugh, 2014b; Lagakos and Waugh, 2013).

Related to structural change, an agricultural intensification literature focuses on the relative scarcity of key production factors and the implications for pathways of agricultural intensification and technical change in a general equilibrium economy (Boserup, 1990; Headey and Jayne, 2014; Ruthenberg, 1971; Hayami and Ruttan, 1971; Block, 2013; Pardey, 2014). These two very rich literatures are connected through labor productivity and its implications for labor allocation across sectors as well as for labors use in agricultural production activities. Agricultural labor productivity is then determined, in equilibrium, by the marginal product of agricultural labor compared to non-agricultural labor, and by the marginal product of labor in agricultural production compared to the marginal product of other agricultural inputs. Labor exits are driven by the equalization of marginal returns across sectors, with a high productivity sector pulling labor from a low productivity sector. Usually, agriculture is the low productivity sector that loses labor over time. However, in an open economy, a country can gain a comparative advantage through agricultural productivity growth, which can lead to an increase in agricultural labor shares (Gollin, Parente, and Rogerson, 2002). These features of structural change and agricultural intensification highlight the interconnectedness of all sectors through factor markets, and the extent to which labor movement across sectors mediates these processes.

Structural change has played out to varying degrees throughout the world, with different countries following many different patterns of growth in labor productivity and exits of labor from agriculture. Much of today's understanding of structural change comes from the experiences of countries that went through the process during the 1960s and 1970s, as adoption of high yielding, modern varieties of rice and wheat overtook Latin America and Asia. Today, most Sub-Saharan African countries are still situated at the very early stages of this transformation.

While the experiences of Asian and Latin American economies are informative, it is important to recognize the different initial conditions that Sub-Saharan African countries face. First, the agroclimatic conditions in Sub-Saharan Africa are very different from those in Asia. African agriculture, like Latin American agriculture, is characterized by upland farming environments, though development of irrigation infrastructure has been incredibly limited in Africa. Second, the dominant non-agricultural sector in Africa is services. Most industry sector activity generates effectively non-tradable goods. Third, most non-farm activities involve self-employment rather than wage employment.

The classic two-sector model with a modern, export-oriented industry sector and a traditional smallholder farmer sector is therefore not immediately relevant in the African setting. Across countries, there tends to be more convergence in industrial sector labor productivity than in agricultural labor productivity (Lagakos and Waugh, 2013; Rodrik, 2014b). Growth in industry is then linked to global competitiveness, posing a challenge for developing country prospects. With open markets, cheaply available manufactured import goods, and global convergence in industry sector productivity, it seems unlikely that industry sector activities will take off (Rodrik, 2014b).

Given differences between Africa and other regions, the role that agricultural productivity growth can play in reducing poverty in smallholder-based African economies today is subject to debate (Collier and Dercon, 2014; World Bank, 2008). Many note that countries with open economies need not rely only on agriculture for labor-intensive (and thus pro-poor) economic growth. While the agricultural sector is likely to play an important role as an engine for economic growth in most poor countries, it is crucial only in certain settings, such as landlocked and resource-poor countries, which often are characterized by their low agricultural potential and/or vulnerability to price collapse during bumper harvests (Dercon, 2009). The difficult irony is that the countries that most need agriculture face the most difficult constraints in achieving agricultural productivity growth.

1.3 Household models of agricultural labor exits

A robust household literature in development economics focuses on modeling occupational choice as it relates to productivity, sector participation, and income growth, the key metrics of structural change (Foster and Rosenzweig, 2007). The micro literature has focused less on sectors in which households participate and more on the modalities by which labor is supplied, distinguishing between self and wage employment, and between formal and informal employment arrangements. Compared to middle and high income countries, the least developed countries, are characterized by the large share of labor supply to agriculture, the importance of unpaid farm labor and self-employment in labor supply, and low investment levels in human capital (Behrman, 1999). African economies, in particular, are characterized by particularly high levels of self-employment and labor market informality.

Early studies of occupational choice in developing countries addressed the ability of subsistence farmers to participate in formal wage labor markets. The literature began with two-sector surplus labor models premised on a formal, urban sector characterized by wage employment and an informal , rural sector characterized by self-employment in subsistence farming (Harris and Todaro, 1970). High rates of urban unemployment and persistence of rural-urban wage discrepancies were explained using a two-sector model with urban employment uncertainty. Within this framework, improving productivity of the subsistence good could slow rates of migration to poor, urban slums.

Another class of studies focuses on explaining employment and productivity patterns using labor market frictions. Workers choose between wage and self- employment in the presence of capital market imperfections (wealth inequality) and information asymmetry (worker monitoring) (Banerjee and Newman, 1993). These market failures explain preference for wage employment by poor workers. In a dynamic setting, higher initial inequality in asset distribution explains a countrys transition towards a manufacturing based economy, as opposed to a cottage-industry based economy characterized by higher self-employment.

As the migration literature that followed has shown, the process of structural transformation necessarily involves some occupational as well as geographic mobility. Though occupational shifts can occur in situ, ultimately people shift both out of agriculture and from rural to urban areas (Collier and Dercon, 2014). Migration out of rural areas and occupational shifts out of agriculture have been associated with poverty reduction in long term panels of rural households (Beegle, Weerdt, and Dercon, 2011).

Household decisions to allocate some labor to non-farm activities have been linked to growth in agriculture and to growth in industrial sectors , as well as to population density and market access (Headey and Jayne, 2014; Haggblade, Hazell, and Reardon, 2007; Reardon et al., 2006). Household income diversification decisions are influenced by growth in the industry and services sectors (Foster and Rosenzweig, 2003; Haggblade, Hazell, and Reardon, 2010), as well as by increased population density and heightened access to population centers (Haggblade, Hazell, and Reardon, 2007). Some households may diversify in order to accumulate assets, while others may do so in order to absorb excess household labor or spread household risk (Barrett, Reardon, and Webb, 2001; Barrett, 2005).

Households may face important barriers in diversifying their occupations or migrating (Barrett, Reardon, and Webb, 2001). In many cases, non-farm work requires quite different skill sets than farm work, and so individuals with farming skill sets must find ways to retool (Rodrik, 2014b). By influencing the ability of individuals to move between sectors, such barriers shape the overall opportunity to achieve labor productivity growth through structural change. Jolliffe (2004) examines the effect of education both on households allocation of labor between the on-farm and off-farm sectors and on the returns to time worked in each sector, lumping together both wage and self-employment non-farm opportunities. In two stages, he estimates the effect of education on income, first through households labor allocation decisions and then through the returns to labor allocated towards each activity. The analysis shows that the returns to education are higher off-farm than on-farm, and that better educated farmers also choose to supply a higher share of their labor off the farm.

1.4 Agricultural technology and investment

While several literatures address occupational choice in developing countries, and several literatures address technology impacts in developing countries, only a few studies have addressed the impacts of agricultural technologies on occupational choice through a structural change framework. Because agricultural productivity growth is central to structural transformation, many have called for government investment in food crop productivity in order to promote economic development (Ranis, 2004; Evenson and Gollin, 2003). The ability to facilitate structural transformation by promoting technology growth in staple food production is therefore of widespread interest.

Eswaran and Kotwal (1985) explore the co-existence of different types of agricultural employment levels in the context of agricultural technical change. In their model, unskilled labor is substitutable for some tasks (e.g., harvest) but not for others (e.g., planting and fertilizer application). Workers then choose between high commitment labor (higher employment level, with repeat employment but a chance of being fired for under-performance) and low commitment labor (lower employment level and no repeat employment) by solving a dynamic optimization problem. They find that, while high commitment labor has a lower equilibrium wage rate than low commitment labor, workers prefer to participate in high commitment labor because it allows them to supply more labor. Labor saving technological change increases overall demand for high commitment workers, while labor using technological change increases demand for low casual farm workers.

Foster and Rosenzweig (2007) examine the relationship between high yielding crop varieties and occupational choice, both for smallholder farmers and for landless laborers. They develop a rural general equilibrium model with three production sectors (local non-tradables, non-local factory production of tradables, and farming). They find that agricultural productivity growth is inversely correlated with growth in non-agricultural wages.

Bustos, Caprettini, and Ponticelli (2012) examine the micro-economic impact of agricultural technology on county level agricultural labor shares in Brazil using census data. They find that release of labor-saving herbicide tolerant soy varieties decreases agricultural employment, while a labor-using technology of maize intensification increases agricultural employment. An ex post impact assessment of agricultural technology adoption of this nature would be premature in the Sub-Saharan African context, where technology adoption levels are much lower.

Technological change is of course induced by factor scarcity. Therefore, it is important to better understand the demand for productivity-enhancing technologies. Demand for technology is often very context specific. It is important to understand where technologies will have highest returns and be demanded by farmers. Policy makers use the incredibly high participation rates in agriculture to justify investing in agriculture. Because returns are heterogeneous, it is important to target investments where the returns will be highest within agriculture, and to avoid investing in agriculture when the returns are lower than they would be outside of it (Dercon, 2013).

1.5 Research contributions

In my first dissertation chapter, I address, from a micro perspective, the topic of cross-sector labor productivity gaps that has been a recent focus of the macroeconomic literature. In a study focused on Ethiopia, Malawi, Tanzania and Uganda, I find that the agricultural sector is not necessarily a bastion of low productivity, as is commonly believed, but rather a large reservoir of underemployed workers. This result emerges when labor inputs are measured using new, high quality micro datasets. While national statistics suggest that workers in these four countries are 6 times more productive outside of agriculture than in it, on an annual basis, using household data I estimate the number is closer to 3.3 times on average. Across the datasets, agricultural workers work fewer hours per year – 700 hours per agricultural worker compared to 1,850 hours per non-agricultural worker. The powerful result is that average labor productivity outside of agriculture is only 1.4 times greater after accounting for different labor inputs. My finding is consistent Lewis original surplus labor hypothesis, where underemployment is disguised in the low productivity sectors (Lewis, 1954a).

This finding has important implications for those interpreting the very large productivity gaps that are measured from national statistics, and helps solve the puzzle of very large productivity gaps that the macro-economic literature struggles to explain (Gollin, Lagakos, and Waugh, 2014b). It highlights the role of underemployment and, in particular, the need to better understand constraints to labor supply in farming. It also highlights the ability to supply more labor as a key motivation for increasing household participation in high-employment activities outside of agriculture, which is different than the productivity-enhancing motivation that is often embraced.

In my second dissertation chapter, I seek to understand how household occupational choices are related to predicted labor productivity and how occupational choices are likely to respond to productivity enhancing interventions. I examine occupational choice from a structural change perspective, incorporating the ideas of informality, self-employment, and occupational diversification. In doing so, I address gaps in understanding about the links between agricultural productivity and the extensive margin of household level agricultural labor supply (Foster and Rosenzweig, 2007). I also address the lively debate in the development economics community about whether countries can or should bypass the agricultural technology-led transformation approach championed by most countries in Asia and Latin America.

In order to better understand the scope for improving welfare through labor productivity growth in agriculture, and the role that shedding labor from agriculture can play in labor productivity growth, I estimate a two stage, structural model of occupational choice. I use a nationally representative dataset from Tanzania to construct sector participation and returns variables. In the first stage, I estimate and impute returns to labor, and in the second stage, I predict occupational choices using an imputed latent income variable. Then, I simulate the effects of different interventions intended to enhance labor productivity and decompose the within-sector and between-sector welfare gains. My results indicate that, while agricultural productivity enhancing investments are associated with gains in labor productivity and overall welfare, they neither speed nor slow the rates at which labor exits farming. Wage and self-employment participation do respond to wage and self-employment productivity shocks, respectively. This finding highlights the importance of enhancing labor productivity within the margins of participation and especially in places with low population density and poor market access, where farming productivity gains are the only ones likely to impact households. One possible explanation for the low responsiveness of participation in farming to productivity-enhancing interventions is underemployment in agriculture, as highlighted in my first dissertation chapter. These results suggest that inter-sectoral diversification of household labor supply is likely to precede labor exits from agriculture.

In my third paper, I explore opportunities to improve the targeting, ex ante, of agricultural productivity enhancing interventions. A key shortcoming of typical priority setting and investment planning efforts is that they do not explicitly account for the uncertainty that decision-makers face. Priority-setting efforts are also typically quite coarse, and do not make use of spatially disaggregated data that are available and could guide targeting. In collaboration with other Cornell researchers, I respond to a request made by the Agricultural Transformation Agency of the Ethiopian Government to help make better use of spatially disaggregated data in the investment priority-setting process.

We conceptualize a robust, ex ante, profitability metric, estimate it using newly available spatially disaggregated data, and present it in an intuitive decision tool aimed at those designing soil health interventions within the Ethiopian government. We find that our recommendations do allow for refined targeting according to user specified profitability criteria. This approach comprises a novel contribution to the priority-setting literature in developing country agriculture, one that has many applications in the broad and growing domain of supporting climate-smart agricultural investments.

Altogether, these papers contribute to a better understanding of agricultures role in economic development in Sub-Saharan Africa. In particular, I highlight the extent to which underemployment explains observed low levels of agricultural productivity in Sub-Saharan Africa. This calls into question the widely held belief that agriculture is inherently less productive than other sectors, and motivates further exploration of what constrains labor demand within the African agricultural sector. This underemployment also has implications for labor exits from agriculture. I find that, while smallholder households are likely to enter into non-farm self and wage employment as the returns to participation increase, they do not typically exit farming in order to participate in these activities. This finding is quite consistent with a world in which underemployment is widespread. It emphasizes the need to consider labor supply diversification in the context of structural transformation, although the structural transformation literature often does not address livelihood diversification. Seasonality of agricultural labor demand is quite central to this challenge of underemployment, as many agricultural tasks require labor at particular times of the year. Because of the very tight link between in-season rainfall and labor demand in rain-fed agricultural systems, it is important to better understand the role of climate uncertainty in determining labor demand and investment returns, in expectation. Flexible tools are needed to support agricultural investment planning in the face of climate uncertainty.

CHAPTER 2

LABOR PRODUCTIVITY AND EMPLOYMENT GAPS IN SUB-SAHARAN AFRICA

2.1 Introduction

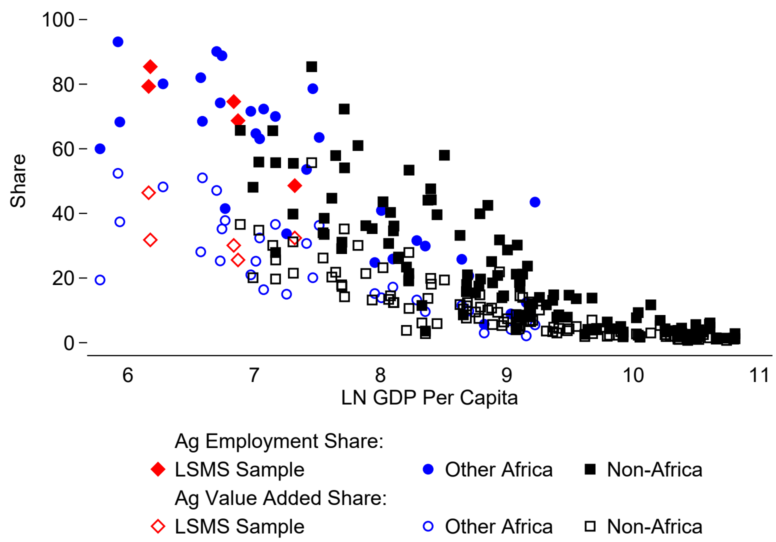
Structural change is integral to economic development. In the development context, it refers both to the reallocation of labor from one low-productivity sector to another, higher-productivity sector, and to the economic growth resulting from this shift. Structural change is a dynamic process powered by several key features — productivity levels within sectors, productivity gaps between them, and the movement of labor from low productivity to high productivity sector(s). The larger the productivity gap between agriculture and other sectors, the larger the opportunity to achieve productivity growth as labor shifts across sectors. In poor economies, agriculture is typically the sector that employs the most people and uses labor least productively. Over time, cross-sector productivity gaps tend to shrink as labor shifts out of agriculture and returns to labor across sectors are equalized through factor markets (Timmer, 1988).

The premise of higher returns to labor outside of agriculture is quite central to structural change. Are these productivity differentials really as high as national accounts data suggest? I use a new micro-level dataset to measure key structural change parameters sector participation, time use, and labor productivity from a micro perspective. This paper draws on the Integrated Surveys on Agriculture from the Living Standards Measurement Study group at the World Bank (LSMS-ISA datasets), which explicitly collect information about respondents time use across sectors. Particular attention is paid to farm labor, which is often neglected in large scale, multi-topic surveys because of the challenges involved in collecting detailed agricultural data. The analysis includes surveys from Ethiopia, Malawi, Tanzania and Uganda.[1] The countries comprising the LSMS-ISA dataset exhibit considerable heterogeneity with respect to GDP per capita, agricultures share of the labor force and economy, and productivity gaps (Figure 2.1).

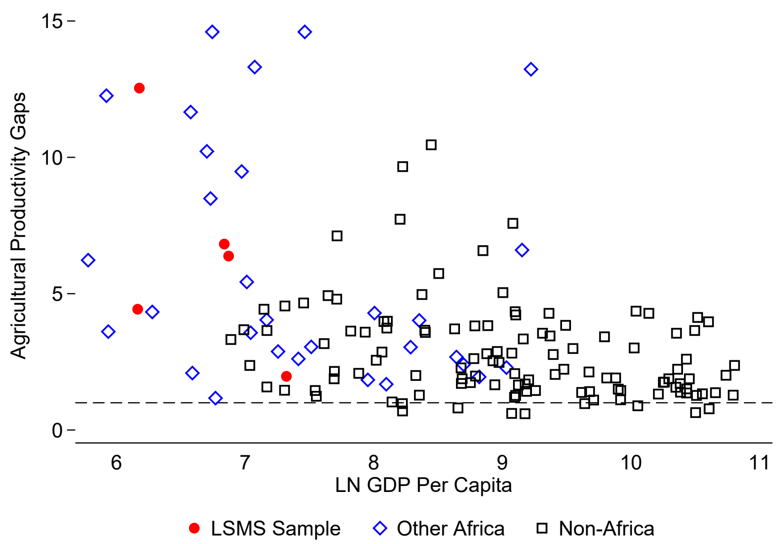
[1]Two other countries, Nigeria and Niger, are part of the LSMS-ISA data collection effort. I have not included Nigeria due to questionnaire incompatibilities with regards to self-employment labor supply. Niger is not included because its rural economy is heavily focused on livestock, and livestock labor inputs are not explicitly collected in the surveys.

Examining productivity gaps from a micro perspective is informative for several reasons. First, individuals and firm owners making labor allocation decisions in developing countries do so based on the micro incentives that they face. Second, micro datasets contain the variables required to address the validity of assumptions that underlie macro statistics. Third, micro datasets allow for productivity measures to be paired with relevant covariates of labor allocation decisions at the household and individual levels. This kind of micro perspective is largely absent from the literature about structural change in African economies. Demographic and Health Survey (DHS) datasets, also micro datasets, are sometimes used to calculate sector labor shares, as an alternative to measures based on population censuses or national accounts (e.g., McMillan and Harttgen, 2014; McMillan and Rodrik, 2011). While DHS surveys have very extensive coverage, they cannot be used to generate measures of labor supply beyond participation, nor can they be used to measure returns to sector participation.

I find that, in four Sub-Saharan African countries, the agricultural sector is not a bastion of low productivity but, rather, a large reservoir of underemployed workers. This result emerges when labor inputs are measured more carefully. Using the LSMS-ISA datasets, I replicate common patterns observed in macro statistics that annual economic output per worker is lower in agriculture than in other sectors and that participation in agriculture is much higher than participation in other sectors. While national statistics suggest that workers in these four countries are 6 times as productive outside of agriculture as in it, I predict the number is closer to 3.4 times on average. This finding is consistent with those of Gollin, Lagakos, and Waugh (2014b), who highlight sources of bias in national accounts measures that lead to under-estimating productivity in agri-



(a)



(b)

Figure 2.1: Panel (a) shows a global cross-section of agricultural labor and employment shares graphed against a log transformation of each country's per capita GDP. Panel (b) shows agricultural labor productivity gaps graphed against the log of GDP per capita (Source: Gollin, Lagakos, and Waugh (2014b)). The horizontal dashed line represents inter-sectoral parity in labor productivity (value = 1).

culture relative to other sectors.

After carefully examining labor inputs, I find that cross-sector productivity gaps observed in national accounts data reflect sectoral differences in employment levels rather than differences in returns per hour worked. Many workers are counted as agricultural because they spend at least some time working on farms. A striking pattern across household surveys is that agricultural workers work fewer hours per year — 700 hours per agricultural worker compared to 1,850 hours per non-agricultural worker. Productivity in agriculture is a lot closer to productivity outside of it when one accounts for systematic differences in labor inputs. On a per-hour basis, labor is only 1.4 times as productive outside of agriculture.

These results suggest that the forces pulling labor into the industry and service sectors may be weaker than is commonly believed. It also casts doubt on the notion that agriculture is intrinsically less productive than other sectors. Because time inputs in agriculture are generally low, possibly due to biophysical constraints, participation outside of agriculture presents the opportunity to supply more hours of labor per year. It is important to better understand the reasons for low labor supply by agricultural workers in order to identify opportunities to increase annual output per agricultural worker.

2.2 Background

This paper focuses on Sub-Saharan Africa, the region with the lowest per capita incomes, largest shares of value added captured by agriculture, largest shares of the work force employed in agriculture, and lowest agricultural labor productivity (Figure 2.1a) (World Bank Group, 2014). According to national accounts data, labor in developing countries is 4.5 times more productive outside of agriculture than in it. In middle income countries, the ratio is 3.4, and in high income countries, it is 2.2. Within African countries, non-agricultural labor is 6 times more productive outside of agriculture than in it[2] (Figure 2.1b) (Gollin, Lagakos, and Waugh, 2014b). Other recent studies confirm that large cross-sector productivity differentials persist in Sub-Saharan African countries (McMillan and Harttgen, 2014; Lele, Agarwal, and Goswami, 2013).

Labor productivity in an economy can be improved either within sectors (e.g., through technological gains and capital accumulation) or structurally (e.g., by shifting labor out of less-productive activities and into more-productive activities).[3] During the 1990s, African labor entered agriculture rather than exiting it, thereby suppressing overall labor productivity growth (McMillan and Rodrik, 2011). Since 2000, labor productivity growth within agriculture has accelerated in Eastern and Southern Africa, and in Nigeria (Pardey, 2014; Block, 2013). When recent labor productivity growth is decomposed into within- and between- sector growth, labor exits from agriculture account for about half of recent overall labor productivity growth in Africa (McMillan and Harttgen, 2014; McMillan, Rodrik, and Verduzco, 2014).

Understanding micro level cross-sector productivity differences, and how they relate to sector allocation decisions, is crucial for understanding the forces that power agricultural labor exits. If productivity gaps are indeed as large as African macro statistics suggest, then one must wonder why so much la-

[2]These ratios were calculated using data from Gollin, Lagakos, and Waugh (2014b) and World Bank classifications of countries by income.

[3]For a description of how labor productivity growth can be decomposed into share-weighted labor productivity growth and productivity-weighted labor shifts across sectors, see McMillan, Rodrik, and Verduzco (2014).

bor remains in rural areas and why rural income diversification remains so low (McMillan and Headey, 2014). One explanation is that, though households may face large productivity gaps, they are not able to diversify because of limited human capital, experience, or financial capital. It is also possible that differences in expected returns between sectors are offset by different levels of risk.

Alternatively, national accounts may mis-measure key components of the productivity equation, namely, labor inputs or returns per worker. After examining many of the assumptions used to measure agricultural labor productivity gaps from national accounts data, Gollin, Lagakos, and Waugh (2014b) find a number of biases that inflate estimates of productivity gaps. These biases arise from the methods used to classify workers as agricultural or non-agricultural, the assumption that workers in each sector work an equal number of hours, and the assumption that workers from each sector have the same levels of human capital.[4] Even after correcting for these biases, the authors find that large productivity gaps remain, with an average corrected productivity gap of 3.3 in Africa.

Another explanation for small micro gaps and large macro gaps is that micro gaps are truly smaller than macro gaps. The cross-sector gaps that households face will be smaller than those suggested by national accounts, should the differential returns to non-agriculture sector activities accrue to owners of capital rather than labor. In capital-intensive industries like mining, wage rates are likely to be much lower than average labor productivity in mining as per national accounts data (McMillan and Harttgen, 2014).

If there is systematic measurement bias across sectors, then productivity

[4]Typically, input quality controls, including human capital, are used in productivity measurement.

gaps calculated from national accounts data will be biased. This paper generates micro-based productivity measures in order to highlight the productivity gaps that households face and to inform the debate about productivity mismeasurement in national accounts data.

Consider the productivity gap between agriculture and services, decomposed into labor inputs (hours per year) and productivity (returns per unit of labor input):

$$GAP_S = \frac{\frac{Y_S}{N_S}}{\frac{Y_A}{N_A}} = \frac{\frac{Y_S}{H_S} * \frac{H_S}{N_S}}{\frac{Y_A}{H_A} * \frac{H_A}{N_A}} = PGAP_S * EGAP_S \tag{2.1}$$

Where Y_S refers to service sector output, N_S refers to the number of service sector workers, and H_S refers to the annual hourly input of a service sector worker. The A subscript refers to the agriculture sector. The gap in annual output per worker between the service and agriculture sectors can then be decomposed into a gap in productivity per hour worked across the sectors ($PGAP_S$) and a gap in employment levels across the two sectors ($EGAP_S$). If cross-sector productivity is equalized in terms of returns to hourly or daily labor at the margin, and labor productivity gaps are largely explained by cross-sector differences in labor inputs, then one can no longer argue that labor is grossly misallocated across sectors even when there is a large cross-sector gap in annual returns per worker. This has implications for the forces that drive labor exits they relate to seeking fuller employment rather than climbing a per-hour productivity gradient.

Claims of underemployment in a smallholder sector have been common, historically. Lewis two-sector model was premised on unlimited supplies of labor, positing labor surplus in subsistence agriculture and also among casual laborers and those self-employed in petty trade (Lewis, 1954b). Lewis also discusses disguised unemployment, whereby many family members supply labor to the household farm, but, should one of the household members be able to find work elsewhere, the same level of output could be maintained if the remaining household members increased their labor supply on the intensive margin. Surplus labor remains relevant, today, in the form of large reservoirs of developing country workers who engage in informal activities and part time work with irregular hours that is characterized by low returns to skill (Gollin, 2014).

Because agricultural labor shares are large in African countries, the potential gains from reallocating labor to higher-productivity sectors are also hypothesized to be large (McMillan and Headey, 2014). A large initial agricultural labor share, rising female education, rising commodity prices, good governance, and agricultural productivity growth all appear to be positively correlated with labor exits from agriculture (McMillan and Harttgen, 2014).

Though African countries seem to be following the same patterns of agricultural labor exits as those followed decades earlier in Asia and Latin America (McMillan and Harttgen, 2014), there are some important differences. The services sector, which is characterized by relatively low productivity in African countries, has been a primary recipient of labor exiting from agriculture (Rodrik, 2014b). In other regions, industrialization has been core to the structural change process. High levels of informality in the industry and services sectors has lowered their average productivity and suppressed the gains to be exploited from agricultural exits. In Vietnam, important productivity gains resulted not only by shifting labor out of agriculture, but also by shifting labor from informal into formal, higher-productivity firms within the industry sector (McCaig and Pavcnik, 2013).

Growth in labor productivity, overall and within agriculture, has been a strong predictor of poverty reduction because of the important linkages between wages, household self-employment, and the real incomes of the poor. Though land productivity growth typically precedes labor productivity growth, the process of agricultural development is thought to begin when output per agricultural worker increases (Timmer 1988). Agricultural labor productivity growth is particularly important because of the direct effects on the many workers who participate in the agricultural sector, and also because of its effects on growth in other sectors (De Janvry and Sadoulet, 2010; Christiaensen, Demery, and Kuhl, 2011).[5]

Labor is one of several important factors in agricultural production, which also relies on land, capital, and other inputs. Where land and capital are scarce (e.g., due to high population pressure or high interest rates, respectively), labor is used more intensively in farming systems. In aggregate, agricultural labor productivity grew slower than agricultural land productivity between 1961 and 2010 in Africa, which implies that African agriculture has intensified with respect to labor (Pardey, 2014). However, Binswanger-Mkhize and Savastano (2014) find that, while population density has increased in rural areas across LSMS-ISA countries, there has been little evidence of Boserupian agricultural intensification with respect to cropping intensity, area farmed, or irrigation.

[5]Non-agricultural growth also contributes to agricultural growth (Irz and Roe, 2005). Empirically, agricultural growth has been shown to contribute more to poverty reduction than non-agricultural growth (Christiaensen, Demery, and Kuhl, 2011).

2.3 Data and variable construction

To examine labor productivity gaps from a micro-economic perspective, I generate labor productivity measures and other key variables from the Living Standards Measurement Survey Integrated Surveys in Agriculture (LSMS-ISA) dataset. I draw on a cross-section of recent LSMS-ISA datasets available, comprised of the Ethiopia *Rural Socioeconomic Survey* (2013-14), the Malawi *Integrated Household Survey* (2010-11), the Tanzania *National Panel Survey* (2010-11), and the Uganda *National Panel Survey* (2010-11). LSMS-ISA surveys were implemented by each countrys national statistics office, with technical support from the World Bank Development Economics Research Group. These datasets are nationally representative, including urban and rural households regardless of occupation or sector of employment. Rural and urban areas are defined by each countrys statistics office.

Table 2.1 depicts the basic characteristics the datasets used in this study. It is worth emphasizing the novelty of LSMS-ISA datasets. Surveys of farming populations often collect detailed plot-level farm management information similar to the LSMS-ISA surveys, but they do not also include information on time use off the farm, and generally do not sample non-farming households. The multi-topic, multi-purpose LSMS-ISA questionnaire includes questions on labor market participation, labor inputs into household farm and non-farm enterprises, and returns to enterprises and labor market participation. They are also internationally comparable to some extent, allowing for cross-country comparisons.

Using the LSMS-ISA data, I construct individual level, annualized labor supply aggregates for three types of activities household operated farm enterprises

Table 2.1: Dataset Characteristics

	ETHIOPIA 2013-14	MALAWI 2010-11	TANZANIA 2010-11	UGANDA 2010-11
Households in sample	5,262	3,247	3,846	2,633
Urban households (share)	0.173	0.245	0.307	0.163
Household size	4.85	4.67	5.09	4.89
(sd)	(2.31)	(2.25)	(2.93)	(2.68)
Household size, adult equiv.	3.94	3.97	4.13	3.70
(sd)	(1.90)	(1.88)	(2.38)	(1.98)
Farm operators, all households (share)	0.772	0.794	0.713	0.790
Farm operators, rural households (share)	0.919	0.943	0.888	0.882
Annual consumption per person, USD PPP, urban HHs	1,600	2,000	2,246	1,641
(sd)	(1,912)	(2,382)	(1,856)	(1,727)
Annual consumption per person, USD PPP, rural HHs	830	748	1,008	675
(sd)	(1,231)	(606)	(820)	(1,040)

(farms), household operated nonfarm enterprises (NFEs), and wage labor market participation. Labor supply recall questions differ in the LSMS-ISA surveys by type of activity. Appendix tables B.1, B.2, and B.3 contain detailed information about the construction of all of the variables used in this analysis.

Wage labor supply variables are generated over a twelve month recall period from individuals reported number of months worked over the last year, typical number of weeks worked per month, and typical number of hours worked per week. In the agriculture modules of the surveys, labor inputs by individual household members are collected for each farm plot. These inputs are aggregated for each household member to generate the annual own farm labor supply variable. For non-farm enterprises, participation by household members is flagged at the firm level. NFE labor supply collection differs slightly from country to country, as detailed in Table B.3 in the appendix.

Systematic measurement error in construction of labor supply variables is particularly concerning, should respondents recall different types of activities with different errors. Differences in recall period (through questionnaire design or timing of interview) or differences in recall ability for different activities (e.g., rare, "salient" events vs. common ones) can lead to differences in household responses (Beegle, Carletto, and Himelein, 2012; Bound, Brown, and Mathiowetz, 2001). The possibility of measurement error in the constructed labor supply aggregates is addressed in Section 2.5 of this paper.

Next, I construct aggregates of labor demanded by household operated farms and NFEs, which include hired labor in addition to labor supplied by family members. Of interest are both the number of firm workers and the total labor inputs supplied by workers to each firm. In the case of farm enterprises, we have a good measure of labor inputs, the number of household members who work on the farm, and the total number of hours worked by household members and hired workers. We do not, however, observe the number of employees hired. It is quite common for farm households to hire in some labor (between 30 % and 94 % of farms do it). In order to avoid under-estimating the total number of farm workers, I predict the number of hired workers by assuming that hired workers work the same hours as own farm workers. In the case of NFEs, we universally observe the number of hired workers but not the hours that they supply to the firm. Non-farm enterprises do not commonly hire workers. In all cases, fewer than 19% of households operating an enterprise hire in any workers.

Returns to labor market participation are comprised of the gross total wages received by wage workers, including in-kind payments (e.g., meals received) and gratuities. Costs of participating in wage labor markets are not measured so it is not possible to construct a net returns measure. The returns to operating a farm enterprise are based on net farm revenue, which is analogous with the "value added" concept that underlies national accounts data. The net value of farm output is derived from the Rural Income Generating Activities (RIGA) calculations and includes the value of own-consumed farm output as measured through the consumption module (Davis et al., 2010). For non-farm enterprises, reported enterprise profit is considered a more reliable measure of net firm revenue than a constructed measure based on gross revenues minus costs (de Mel, McKenzie, and Woodruff, 2009). Where available, I construct the annualized firm level net revenue variable using reported profits. Otherwise, I use the household estimate of gross NFE revenue and subtract household estimated costs. To facilitate cross-country comparison, all measures of returns are converted to constant international dollars using the purchasing power parity conversion factor for private consumption from the World Banks World Development Indicators.

Using the labor supply variables and the returns variables, I construct average labor productivity variables. These are done separately for the three types of activities wage labor, farms, and NFEs as a simple ratio between returns to an activity and labor inputs into the activity. Two types of average labor productivity measures are constructed. The per-worker measure is based on output per worker per year. The per-hour measure is based on output per hour of labor supplied to each activity per year. Because we do not observe how many hours hired workers supply to NFEs, I am unable to generate per-hour productivity measures for these firms.

The next task involves generating sector level labor productivity measures, which aggregate, at the sector level, returns from and labor inputs to self-employment, wage employment, and farming. First, all activities are assigned to their respective sectors of the economy (i.e., agriculture, industry, or services). Following McMillan and Harttgen (2014), I group these into the general categories of agriculture (primary agricultural, livestock, and fishery and forestry production), industry (manufacturing, mining, construction, and public utilities), and services (wholesale and retail trade, transport and communication, finance and business services, and community, social, personal and government services). I generate sector level aggregates of labor supply and returns for each household. Farm activities are classified as agricultural. Wage labor and NFE activities are classified using the Industry Standard Industrial Classification (ISIC) codes provided with each activitys description. An additional sector definition of unknown is used when individuals report jobs for which no description or sector code is available. These labor sources most likely occur in the agriculture sector, but I avoid assuming so.

The hourly agricultural labor supply aggregates do not include livestock and post-harvest labor. And the corresponding agricultural labor productivity measures do not include revenue from livestock in the numerator. In the per-person agricultural labor productivity measure, the numerator includes net livestock revenue (taken from the RIGA dataset), and the denominator includes workers who participate in livestock rearing. Table 2.2 presents a high level overview of the contents of each constructed productivity variable.

2.4 Corroborating Macro and Micro Evidence

2.4.1 Sector Labor Shares

Often in the macro measures of sector productivity, individuals are constrained to one sector of participation, and it is assumed that individuals in each sector work the same number of hours and do not supply labor to secondary sectors. Usually, each sectors labor inputs are assumed to be of the same skill and not adjusted for different levels of human capital. Initial examination of these assumptions using LSMS-ISA data suggests that they are indeed problematic and lead one to overestimate labor supplied to agriculture relative to other sectors, thereby artificially inflating estimates of the labor productivity gap between agriculture and other sectors (Gollin, Lagakos, and Waugh, 2014b).

Figure 2.2 depicts three different measures of sector labor shares constructed using LSMS-ISA data along with two other commonly used measures from national accounts and from Demographic and Health Surveys (DHS).[6] The first

[6]The national account measure comes from the sector employment dataset published by the International Labor Organization (ILO) and sector value added measures from the United Nations (UN) National Accounts Statistics, accessed through the World Bank World Development Indicators database. Sector employment statistics are generated from population censuses or labor force surveys, with methodologies varying across countries. The Malawi agricultural labor

Table 2.2: Overview of productivity variable construction

	Per Person	Per Hour
Activity level productivity measures		
Farming	Farm net revenue / (# own farm workers + predicted # hired in farm workers)	Farm net revenue / (hours worked by own farm and hired in workers)
Self Employment	Firm profits / (# HH firm workers + # hired in firm workers)	*not generated*
Wage Employment	HH wage returns / # HH members participating in wage employment	HH wage returns / # hours worked for wages by HH members
Sector level productivity measures		
Agriculture	(HH net returns to farming + livestock + hired out ag wage labor) / (# hh members who participate primarily in ag sector)	(HH net returns to farming + hired out ag wage labor) / (hours worked on own farm + hours hired out for wages in ag)
Industry	(HH net returns to ind sector NFE + hired out ind sector wage labor) / (# hh members who participate primarily in ind sector)	(HH net returns to ind sector NFE + hired out ind sector wage labor) / (hours worked on own ind sector NFE + hours hired out for wages in ind sector)
Services	(HH net returns to ser sector NFE + hired out ser sector wage labor) / (# hh members who participate primarily in ser sector)	(HH net returns to ser sector NFE + hired out ser sector wage labor) / (hours worked on own ser sector NFE + hours hired out for wages in ser sector)

column in Figure 2.2 is based on the labor supplied by all adult individuals in the LSMS-ISA dataset.[7] The second is based on the primary sector of each adult individual in the household, i.e., the sector to which each individual supplies the most hours.[8] The third is based on the primary sector of the household head. This sub-sample includes individuals who reported positive hours worked in any sector.

Several patterns are common to all of the countries depicted in Figure 2.2. First, agriculture is the dominant sector of participation across all data sources and aggregation methods, and participation in services is generally more common than participation in industry. Second, agricultural labor share estimates are slightly higher when they are based on all adult individuals in a household rather than just the household head. This suggests that household non-heads are more likely to work in agriculture than household heads. Third, hours-based agricultural labor shares are lower than participation-based shares, which is further explored below. Fourth, individual-based estimates of agricultural labor shares are lower than national-accounts measures in all countries. Fifth, the DHS-based measures of agricultural participation shares are quite a bit lower than the LSMS-based individual participation shares in Ethiopia and Malawi. This implies that, in these countries, DHS-based labor share estimates might under-estimate agricultural labor shares and therefore overestimate labor productivity in agriculture relative to other sectors. Since individuals self-identify their primary sector in DHS surveys, it is possible that respondents involved in

---

share estimate is from Gollin, Lagakos, and Waugh (2014b). The DHS measure of sector labor shares are based on the self-reported primary occupations of adult respondents who work and do not attend school. These are taken from McMillan and Harttgen (2014).

[7]Following McMillan and Harttgen (2014), adulthood is assumed to begin at age 25 to avoid confounding labor shares and educational attainment. Labor shares are robust to the adulthood threshold used.

[8]In most LSMS-ISA surveys, respondents are not asked to name their primary occupations explicitly.

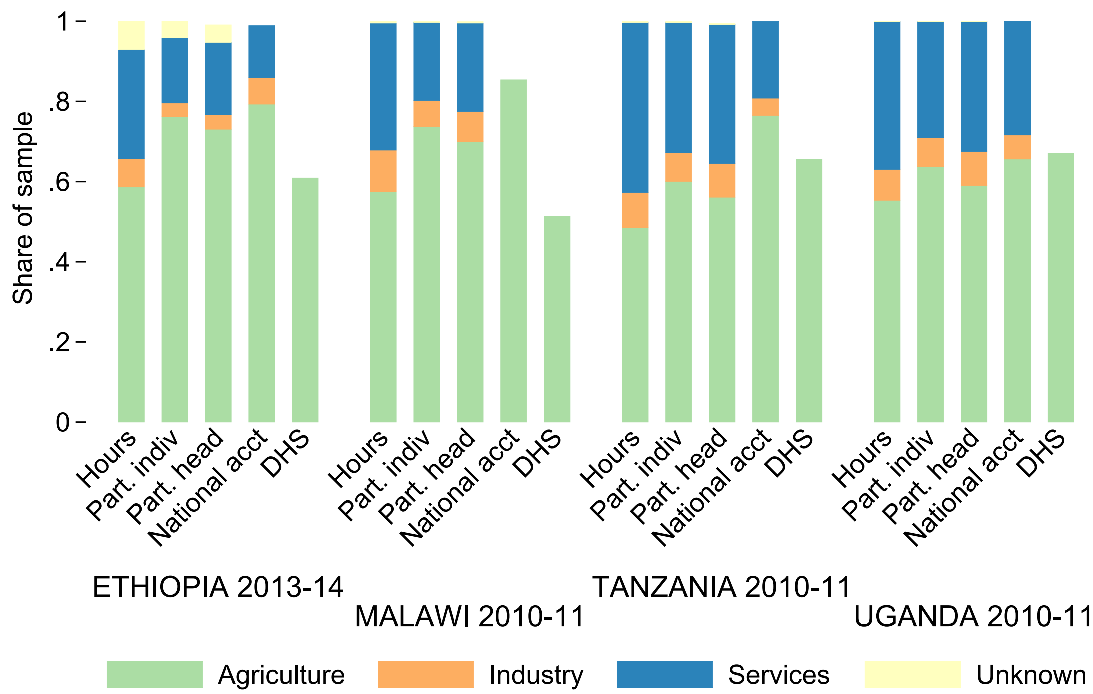


Figure 2.2: Comparison between different estimates of sector labor shares. The "Hours" measure is from variables generated using LSMS-ISA data. The "Part. indiv" measure is based on the primary occupation (most reported hours) of individuals in the dataset. The "Part. head" measure is based on the primary occupation of the household head. The "National account" measure is from the World Development Indicators database, and the "DHS" measure is based on DHS surveys, as described in the text.

multiple sectors are more likely to identify the non-agriculture occupation even though it accounts for a lower share of labor supplied.

Per-person productivity measures based on categorizing individuals by their primary sector of occupation implicitly ignore individuals contributions to secondary sectors. They also assume that participants in different sectors supply equal hours of labor. Both assumptions are problematic when individuals supply labor to secondary sectors, or when there are systematic cross-sector differences in hours supplied. Indeed, LSMS-ISA datasets suggest that both assumptions are violated.

Figure 2.3 examines the one sector assumption, categorizing individuals by their primary sectors and depicting the average hours supplied to individuals primary as well as secondary sectors. The data imply that both the equal-hours and primary-sector assumptions are problematic. While those who are primarily categorized as agricultural laborers do not supply much labor to other sectors, workers who are primarily in industry or services sectors do supply labor to agriculture. Because secondary sectors are an important part of individuals labor supply, we likely underestimate labor supplied to agriculture by ignoring the labor supplied by individuals who participate in agriculture as a secondary activity, thus leading to an overestimation of labor productivity in agriculture relative to other sectors. Gollin, Lagakos, and Waugh (2014b, 's) analysis on secondary sector bias suggests that labor supply to non-agriculture by agriculture workers is greater than labor supply to agriculture by non-agriculture workers. These data indicate bias working in the opposite direction, with non-agricultural workers supplying more agricultural labor than agricultural workers supply to non-agriculture.

Violation of the equal hours assumption, on the other hand, leads to overestimation of agricultural labor inputs. Figure 2.4 depicts the average hours worked in a sector by those who participate in it. Generally, those working in non-agricultural sectors supply significantly more hours than those working in agriculture. Gollin, Lagakos, and Waugh (2014b) address the differences in hours supplied by agriculture and non-agriculture workers, using rural and urban distinctions where sector distinctions are not available. They find that,

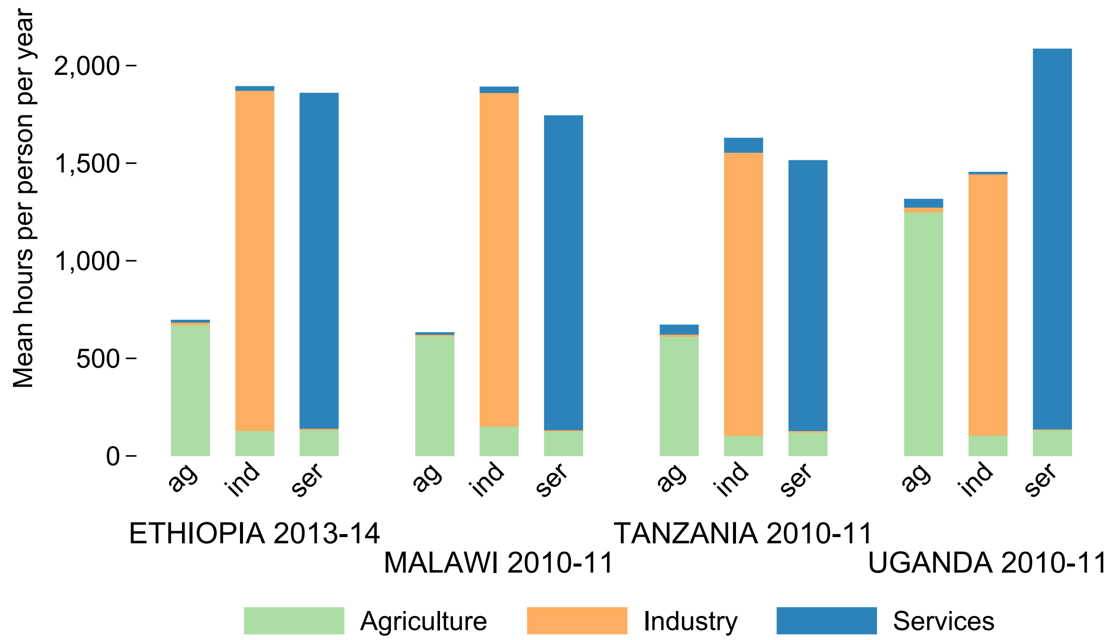


Figure 2.3: Average hours supplied by individuals to all sectors, categorized by each individual's primary sector of participation.

in poor countries, non-agricultural workers supply 1.3 times as many hours as agricultural workers to their respective sectors. This analysis confirms higher supply of labor to non-agriculture by non-agriculture workers than supply of labor to agriculture by agriculture workers, though our cross-sector differences in labor supply are large (between 2.3 and 2.5 in Malawi vs. Gollins 1.45, between 2.4 and 2.6 in Ethiopia, and between 2.1 and 2.2 in Tanzania). Our Uganda estimates, however, are smaller (between 1.0 and 1.6 vs. Gollins 2.3). These ratios are based on any form of sector participation (primary or secondary). When the sample is restricted to individuals who primarily participate in each sector, the ratios are quite similar.

Overall, the LSMS-ISA datasets suggest large gaps between hours supplied

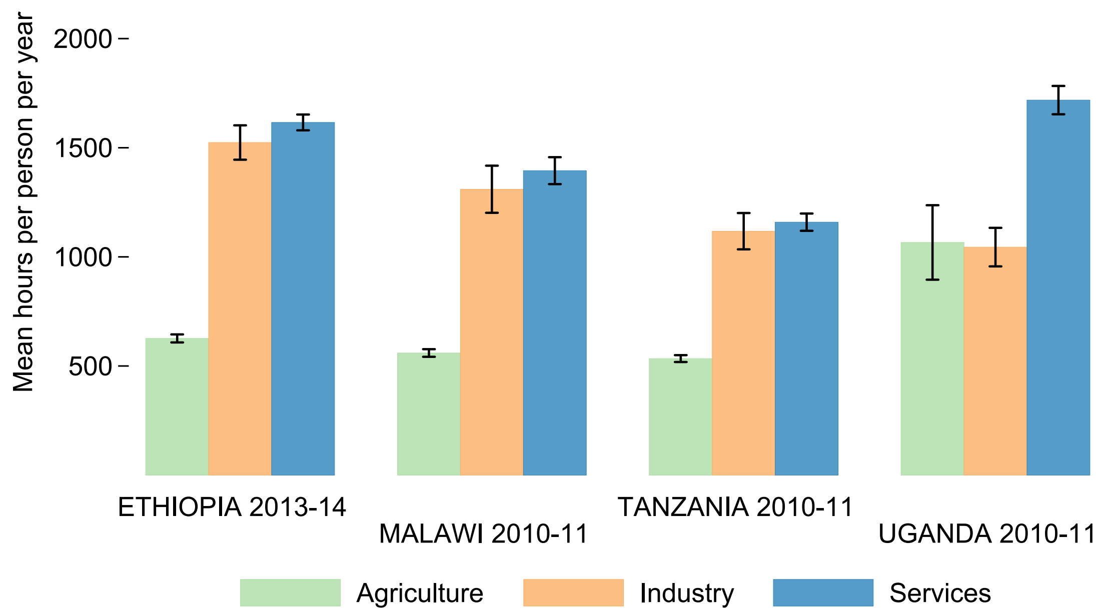


Figure 2.4: Average hours worked per year by sector participants. This sample includes all individuals between the ages of 16 and 65 who actively participate in the labor force. 95% confidence intervals for the mean are also depicted.

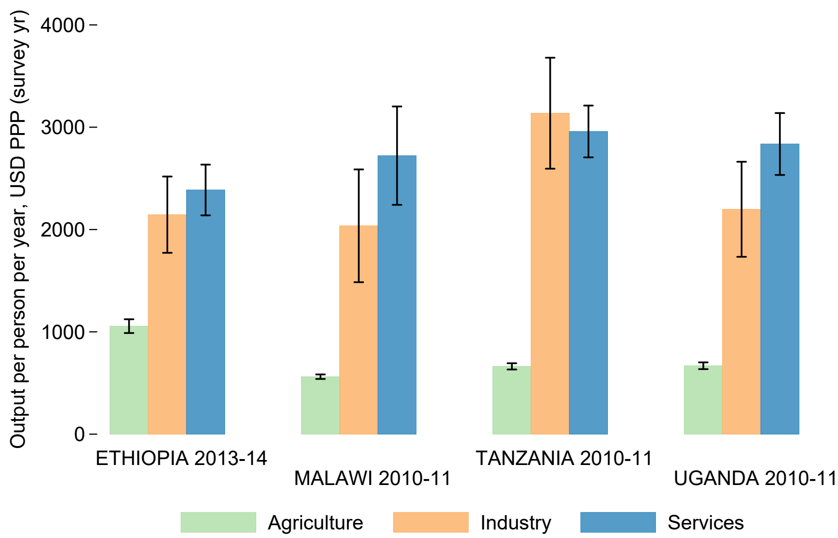
to agriculture and hours to industry and service sectors. By calculating sector labor inputs based on participation rather than hours worked, one over-estimates labor inputs in agriculture compared with other sectors.

The bars labeled Hours in Figure 2.2 show the net effect of the equal hours assumption and the no-secondary-sector assumption on labor share measurement bias. In this case, the sources of bias offset each other. In all countries, agricultures share in labor is lower when an hour-based measure is used than when the LSMS-ISA participation-based measure is used. These results suggest that agricultural productivity may be underestimated relative to other sectors when participation-based labor shares are used. When the intensive margin of labor supply is controlled for by using hours-based labor share measures, estimates of agricultural productivity are relatively higher, and estimates of productivity gaps are smaller. This bias proves extremely important to any discussion about structural change in Sub-Saharan Africa.

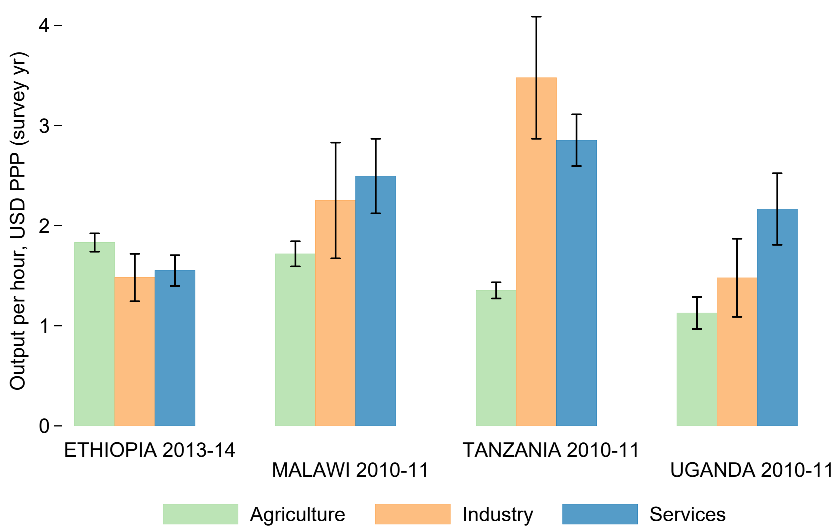
2.4.2 Sector Productivity Gaps

Cross-sector productivity gaps calculated from the LSMS-ISA datasets are indicative of the average productivity differentials that households face when allocating their labor. Figure 2.5a depicts sector-level productivity measures with 95% confidence intervals in four LSMS-ISA countries, based on output per person per year. Output per worker per year is highest in the industry and service sectors, between $2,000 and $3,200 (USD ppp) per worker per year. Agricultural output per worker is between $560 and $1,060 (USD ppp) per worker per year in all countries.

Figure 2.6a depicts micro-level productivity gaps (simple ratios between each sectors productivity and productivity in the agricultural sector) along with national accounts based measures of productivity gaps, gathered for the purpose of comparison. These per worker measures of average labor productivity are not meant to replicate the output per worker measures generated from national accounts, which use different sampling approaches. Corporations are not sampled in the LSMS-ISA surveys, for example, so their activities are only detected through wages paid to workers hired by such firms. Should non-agriculture activities be more capital intensive, then capital ownership differences could explain why macro level productivity gaps are slightly larger than



(a)



(b)

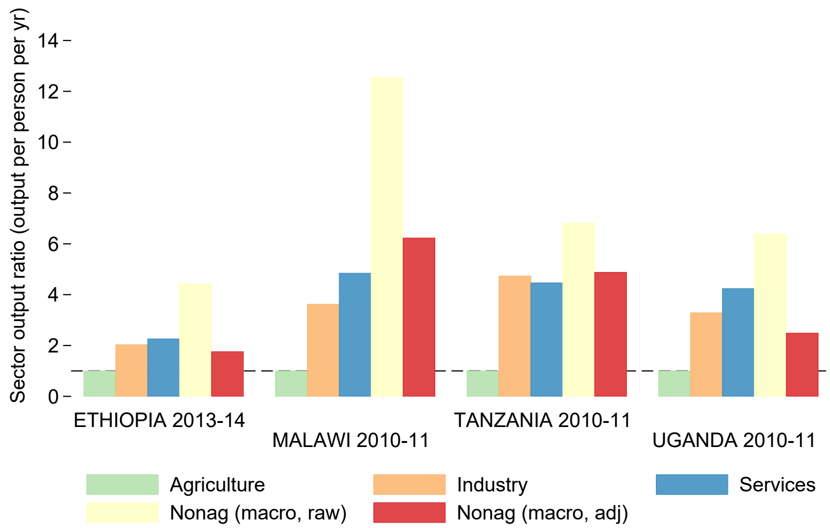
Figure 2.5: Productivity by sector. Panel (a) shows annual value of output per sector primary participant per year. Panel (b) shows output per hour worked per year.

micro level gaps. Gaps in output per worker per year are smaller than national accounts gaps in all countries 2.6a.

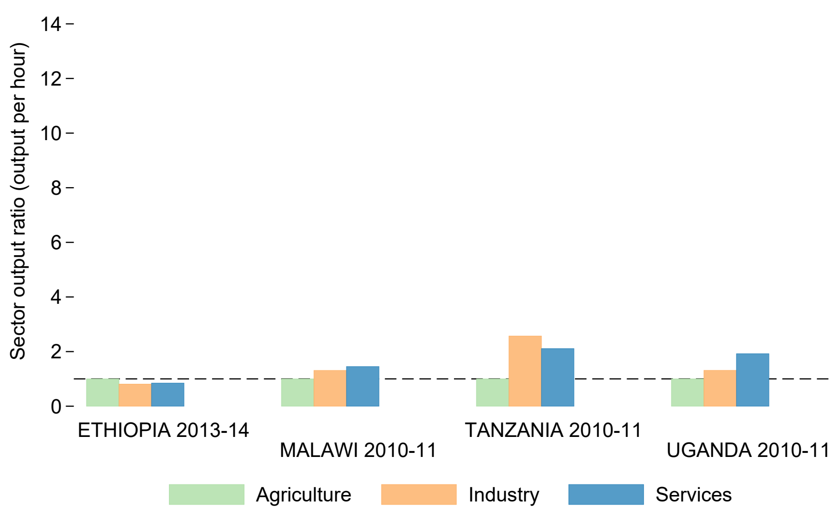
Figure 2.5b shows sector level output per hour of labor worked in each sector. After adjusting for labor inputs (hours worked), returns per hour of labor supplied are between \$1 and \$3.50 (USD ppp) in all sectors. When considering time inputs in each sector, cross-sector gaps in productivity shrink considerably (Figure 2.6b). The hours-based gap measures are much smaller than the per-person-per-year gap measures in all countries. An hour worked outside of agriculture is 0.9 times as productive as an hour worked in agriculture in Ethiopia, 1.4 times as productive in Malawi, 2.1 times as productive in Tanzania, and 1.9 times as productive in Uganda.

Much of the productivity differences observed in national accounts statistics may then be attributable to differences in hours supplied by workers in each sector rather than differences in output produced per hour worked in each sector. Table 2.3 shows each countrys overall gap in output per worker per year, along with the two components of this gap output per hour worked and hours worked per year. Employment gaps explain about half of overall micro level productivity gaps in Uganda, and a larger share in all other countries.

After further disaggregating returns to labor between self-employment and wage employment, it is clear that wage employment brings higher annual returns to participants than does self-employment. Figure 7 depicts productivity gaps at the sector-activity level, with farming as the comparison activity. The sector-activities compared include household-operated farms, household-operated non-farm enterprises (NFEs) in all sectors, and wage labor in all sectors. Because hours or days of labor supplied by hired workers to NFEs are



(a)



(b)

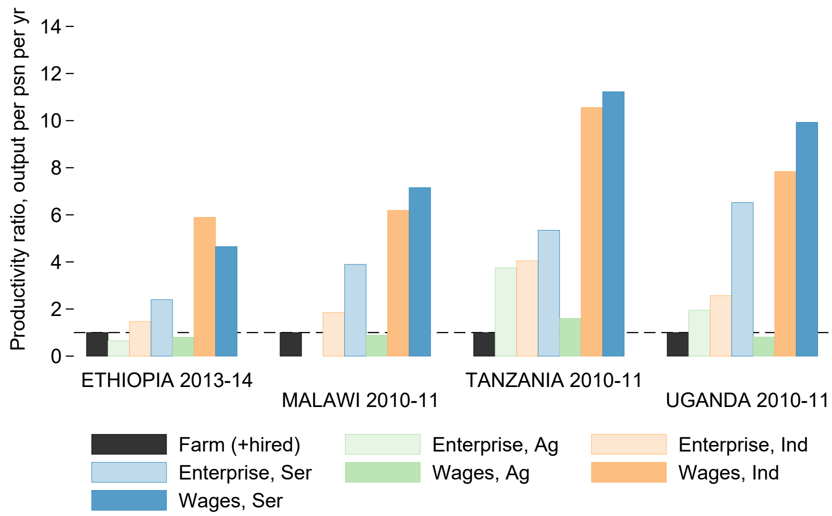
Figure 2.6: Productivity gaps by sector. Panel (a) shows the ratio between productivity in each sector and agriculture based on per-person-per-year productivity measures. The fourth column depicts the raw productivity gaps between agriculture and non-agriculture as constructed using national accounts data, and the fifth column refers to adjusted gaps constructed by (Gollin, Lagakos, and Waugh, 2014b). Panel (b) shows the ratio between productivity in agriculture and in other sectors based on output per time input.

Table 2.3: Ratios between non-agriculture and agriculture in output per worker per year (productivity gaps), hours worked per year (employment gaps) and output per hour worked (per-hour productivity gaps).

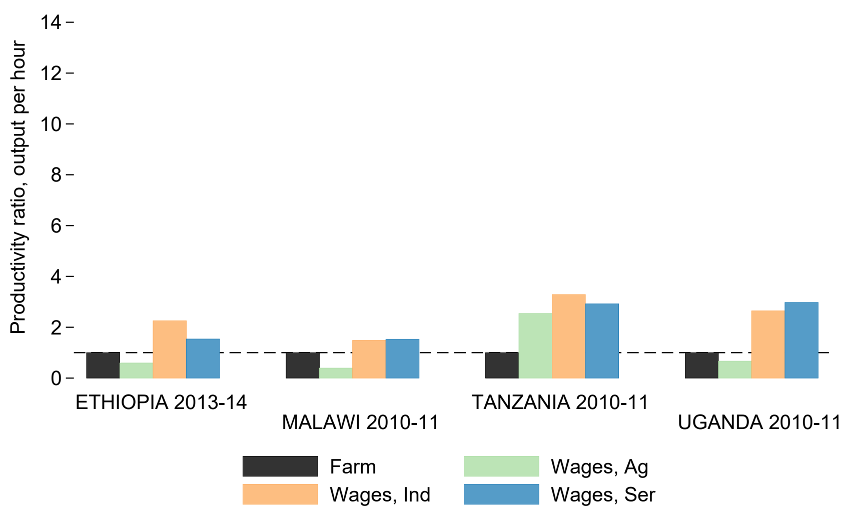
	Per-person Productivity Gaps	Employment Gaps	Per-hour Productivity Gaps
Ethiopia 2013-14	2.25	2.66	0.85
Malawi 2010-11	4.76	3.30	1.44
Uganda 2010-11	4.48	2.10	2.13
Tanzania 2010-11	4.20	2.22	1.90

not collected anywhere besides Malawi, per-hour firm level productivity estimates are not included for NFEs. Wage labor returns should not be interpreted as measures of productivity, especially in the presence of market frictions, of which the evidence is strongly suggestive (Dillon and Barrett, 2014). They do offer a lower bound on the marginal revenue product of rented out labor, and they also provide a benchmark against which individuals in an economy can compare returns to self vs. own employment.

The sector-activity patterns depicted in Figure 2.7a are similar to the patterns observed at the sector level (Figure 2.6a). First, mean returns per participant per year are higher in industry and service sectors than in farming, whether the labor is supplied to NFEs or to wage employment. Within each sector, wage labor brings higher returns per worker per year than does self-employment. Wage laborers in the agricultural sector earn lower annual returns than industry and service sector laborers in all countries. The returns to agricultural wage labor are lower than the returns to own-farm labor in Ethiopia, Malawi and Uganda, and only slightly higher in Tanzania. At the hourly level, productivity gaps between farming and wage employment in industry and service sectors



(a)



(b)

Figure 2.7: Productivity gaps by activity for all households (ratio between mean values for each activity). Panel (a) depicts the ratio between mean farm labor productivity per person per year and the mean labor productivity of other activities (i.e. NFEs and wage labor in different sectors). Panel (b) depicts per-hour productivity gaps for the same activities.

shrink considerably due to differences in labor supply between farm workers and non-agricultural wage laborers.

The existence of cross-sector gaps in output per worker per year suggests there are some forces enticing smallholder farmers into industry and service sector activities. Participation in industry and service sector activities may allow for fuller levels of employment in terms of hours of labor supplied per year. It is not possible, with cross-sectional data, to determine whether agricultural workers tend to work fewer hours because of constraints to labor supply or to labor demand. Biophysical and agronomic characteristics could limit the periods during the year in which farm labor can be used productively. In this case, it might not be possible for individuals to increase their agricultural sector returns by supplying more labor to their farms. Presumably, because labor supply is so low across households and countries, low demand for labor by agriculture is a key constraint. Agricultures role as a low entry barrier sector could help explain both high levels of participation in farming and low per-worker labor supply. Though individuals may aspire, and even attempt, to participate in non-farm activities, they may still return to farming as the sector that can basically guarantee employment. Labor transitions back into agriculture by individuals who had exited farming has occurred in Uganda (Christiaensen and Kaminski 2015). Understanding what limits supply of and/or demand for labor in the agricultural sector is an important topic that is left for future research.

Within urban areas, self-employment in the service sector does not seem to serve as a sink for underemployment as does agriculture in rural areas. One might expect high rates of declaring self-employment due to possibly lower entry barriers than wage employment. Using the LSMS-ISA datasets, Nagler and Naudé (2014) show that self-employment participation correlates include wealth, credit access, and education. Self-employed workers in the industry and service sectors in urban areas tend to supply far more hours per year than do urban wage workers. The annual returns per worker to industry and service sector self-employment are much higher in urban areas than in rural areas, a finding consistent with Nagler and Naudé (2014). By assuming household firms do not hire in outside labor, one can estimate an upper bound on hourly returns to self-employment. These productivity estimates are very low, suggesting workers have a desire to supply labor even despite low returns.

2.5 Robustness of Productivity Gap Measurement

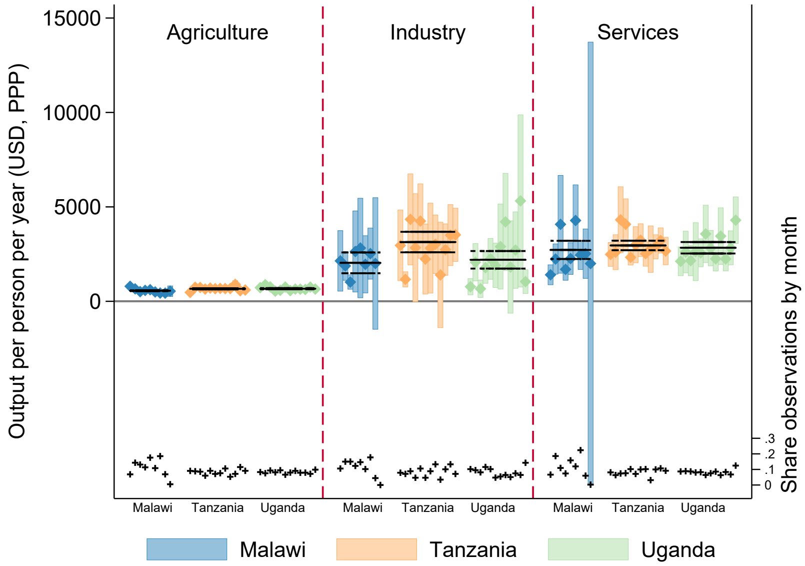
Next, I turn to showing that these productivity gap measures are robust. I am concerned with both the measurement of labor inputs and the returns to labor. The first major concern is sensitivity of labor productivity measurement to survey timing. Labor supply varies seasonally and is elicited over discrete recall periods, raising the possibility that seasonal bias enters into labor productivity measurement. The second major concern arises from questionnaire design issues. Different types of labor supply are collected through different survey modules and elicited in different ways. The goal is to show that the key insights regarding sector participation, labor supply, and productivity are fairly robust to survey design. At the end of this section, I turn to measurement of the returns to labor, showing that the same patterns hold if consumption is used instead of income as a measure of returns to labor.

Annualized labor supply measures of participation and hours worked comprise the denominators of per-person-per-year and per-hour productivity measures, respectively. These aggregates are constructed from more detailed labor supply questions asked of respondents, such as the number of hours worked in the last week, or, in some cases, the number of hours worked in a typical week. One would expect these aggregates to move seasonally due to seasonal patterns in labor supply or a combination of seasonality of labor supply and recall bias in the case of a typical week recall approach.

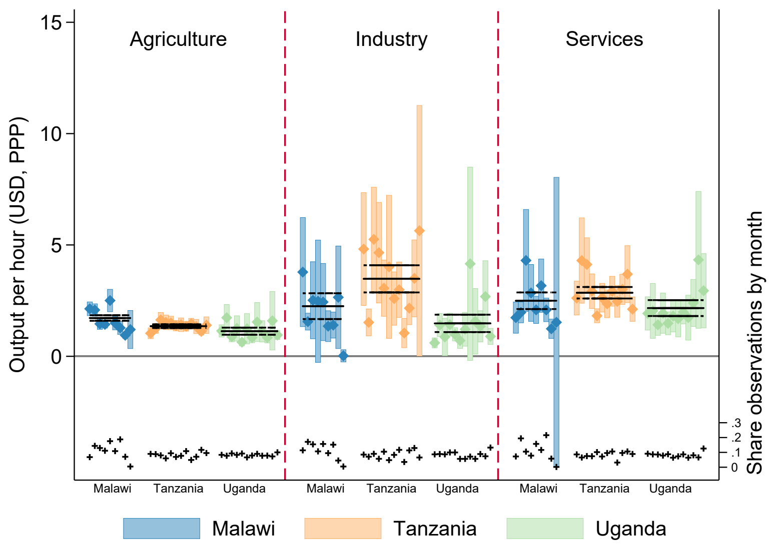
I demonstrate how the per-worker-per-year and per-hour labor productivity measures vary by month of survey visit in Figure 2.8.[9] Each diamond represents a monthly mean productivity measure, and the bar it sits within depicts 95% confidence intervals for the mean. The horizontal solid line represents the annual survey-weighted average for the survey, along with dashed lines above and below representing its 95% confidence intervals. If more surveys are conducted during high or low productivity times within the year, then annual productivity aggregates would be biased. This is especially concerning if different sectors have different seasonality patterns within a country. According to Figure 2.8, there are some months with especially high or low productivity measures, but there does not seem to be a major pattern of over- or under- representing these months.

In order to address concerns that survey timing is somehow correlated with seasonal productivity patterns, I generate new population-month weights to create annually representative measures of per-person-per-year and per-hour productivity for each sector. Using the weights, I also generate annually representative measures of the different components of labor supply. On the ex-

[9] Ethiopia is omitted from this analysis because the questionnaire administration is highly concentrated in a two-month period, with very low coverage in other months.



(a)



(b)

Figure 2.8: Panel (a) shows average annualized output per worker per year by month of household interview (and the 95 % confidence interval for each productivity measure). The horizontal line shows the annual mean for each productivity measure, with the dashed lines above and below depicting their 95 % confidence intervals. The share of observations per month is plotted at the bottom of the figure along the right hand axis. Panel (b) shows sectoral output per hour of labor supplied, along with the annual mean for output per hour worked.

tensive margin, this includes participation on an annual basis. On the intensive margin, this includes participation in the last week conditional on participation in the last year and hours of labor supplied per week. I conduct a t-test for difference between the survey weighted means depicted in Section 2.3 and these survey-month weighted means. In Uganda and Tanzania, I cannot reject the null hypothesis of equal means between survey weighted and seasonally corrected measures of productivity or labor supply. In Malawi, there is evidence that, by not correcting for seasonality, agricultural labor supply is under-estimated and wage labor supply is over-estimated. If these biases were to be removed, per person productivity gaps would be the same but per hour productivity gaps would be slightly larger. The effect is small in magnitude (7% of the uncorrected per-hour productivity measure), and the difference is significant at the 10% level. This analysis suggests that seasonal bias due to survey timing does not bias the key labor supply or productivity variables.

Because labor supply variables for different activities are constructed from different types of survey questions, there is concern that differences in labor supply across activities could arise from different survey recall approaches rather than actual labor supply differences. In particular, downward bias in the measurement of agricultural labor supply or upward bias in self-employment or wage employment labor supply would undermine the agricultural underemployment findings. In a recent methodological experiment designed to compare different approaches to measuring farm labor inputs, Arthi et al. (2016) find that end-of-season plot based recall measures inflate farm labor supply considerably. The labor supply aggregates generated using the LSMS-ISA approach are about twice as large as labor supply aggregates generated by weekly eliciting the data from respondents in person or over the telephone, which is considered to be a more accurate approach. The LSMS-ISA approach also generated aggregates that were larger than those generated using a standard, stylized seasonal recall of days worked, without collecting plot specific information. These findings suggest that, given survey design, labor supply for smallholders is likely to be over-estimated rather than under-estimated. If this is the case, then underemployment within agriculture would likely explain an even larger share of productivity gaps. There has been little research on recall bias for wage and self-employment labor in the developing country context. However, the experimental evidence suggests that, in the LSMS-ISA surveys, farm labor aggregates are higher than they would be if measured using a stylized seasonal recall approach, which is comparable to the approach used to gather wage and self-employment data in the LSMS-ISA datasets.

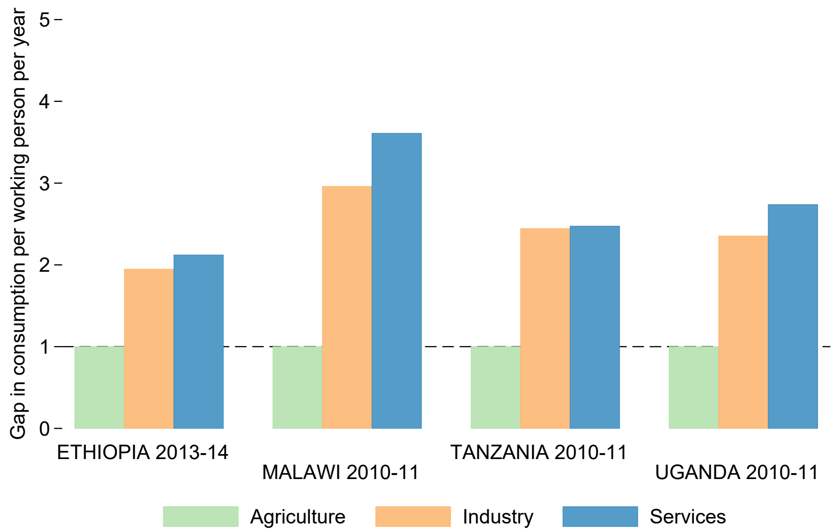
Sector productivity gap estimates are sensitive to prospective measurement error of labor returns farm and firm revenue and wage earnings. I use an alternate measure of returns to labor, household consumption, to ensure that the measurement of productivity gaps is robust to measurement of returns. Consumption can be thought of as household profits after participation costs (for wage labor) and production costs (for firms), assuming no savings or dissavings. Because households who face stochastic income generally smooth their consumption from year to year, consumption can be a good measure of permanent income (Bhalla, 1978). It is a central focus of LSMS-ISA surveys to generate consumption aggregates, so this variable plays to the strengths of the data. Consumption aggregates are generated by each countrys statistics office and released with the datasets. They include cash expenditures as well as the imputed value of items that are produced and consumed by the household, such as agricultural goods.

The consumption gap estimates are a ratio in annual per-worker consumption between households participating primarily in agriculture and those participating primarily in industry and services, respectively (Figure 2.9a). I also create an analogous per-hour measure, which is based on consumption per hour of labor supplied by households, including labor supply to secondary sectors. Households are classified by their primary sector of participation. The per-hour measures are shown in Figure 2.9b. These consumption gaps are fairly similar across countries, and are quite similar in magnitude to per-person-per-year productivity gaps. Households primarily in the industry sector consume 2-3 times more per capita per year than agricultural households. Households primarily in the services sector consume 2-4 times more per capita per year than agricultural households. As with productivity gaps, consumption gaps also disappear almost entirely when they are expressed per hour of labor supplied by each household. This suggests that differences in consumption across sectors (as with differences in returns to sector participation) can be explained in large part by differences in hours worked across sectors.

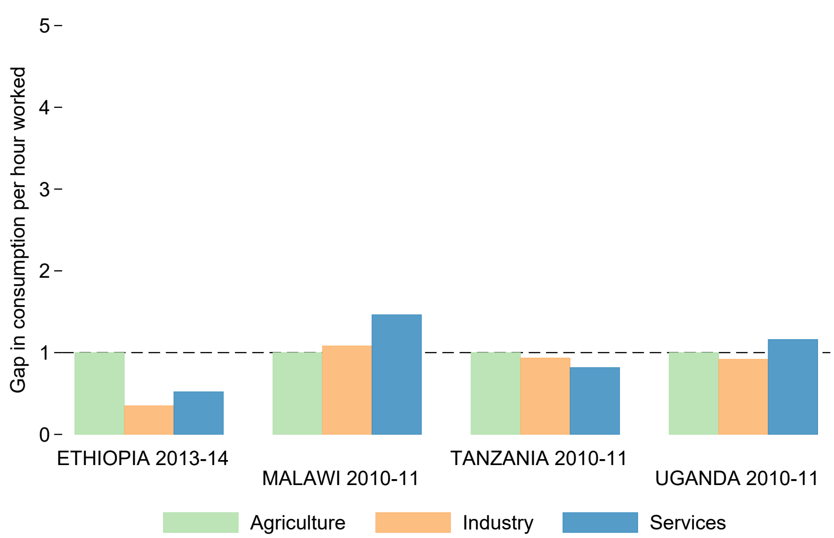
2.6 Exploring the Non-Farm Economy

It is important to understand not only sectors of employment, but also the modalities by which workers supply their labor. Because of growth linkages between agriculture and non-agriculture, the specific types of activities to which workers supply labor can inform the scope for growth linkages between different sectors of the economy.

Recent micro evidence suggests that, while non-agricultural sources of in-



(a)



(b)

Figure 2.9: Panel (a) shows average annualized output per worker per year by month of household interview (and the 95 % confidence interval for each productivity measure). The horizontal line shows the annual mean for each productivity measure, with the dashed lines above and below depicting their 95 % confidence intervals. And the share of observations per month is plotted at the bottom of the figure along the right hand axis. Panel (b) shows sectoral output per hour of labor supplied, along with the annual mean for output per hour worked.

come bring the highest returns across the welfare distribution, the majority of households in African rural areas remain specialized in agricultural income earning activities (Davis, Di Giuseppe, and Zezza, 2014). After controlling for per capita income, though, households in Sub-Saharan Africa have similar diversification levels as households in other regions of the world.

A close examination of the non-farm activities in which households are involved suggests some clear patterns across countries. Workers outside of agriculture are more educated, younger, and less female than agricultural workers. Rural non-farm activities tend to be closely related to agriculture, with strong producer and consumer linkages.

2.6.1 Household and worker characteristics

It is important to explore any systematic differences in characteristics of sector participants, so that they can be taken into account when interpreting sector labor supply data generated from national accounts. Indeed, the macro-economic literature is concerned with systematic differences in human capital across sectors and the implications for bias in productivity measures (Vollrath, 2014).

Industry and service sector workers tend to be younger, on average, than agricultural workers. In all countries, the agricultural work force contains a larger share of women than men, while industry and service sector work forces contain more men than women. Palacios-Lopez, Christiaensen, and Kilic (2015) provide analysis of gender share of agricultural labor supply using LSMS-ISA data, pooling own and hired farm labor supply by gender. They find that women do not necessarily contribute a larger share of agricultural labor, in

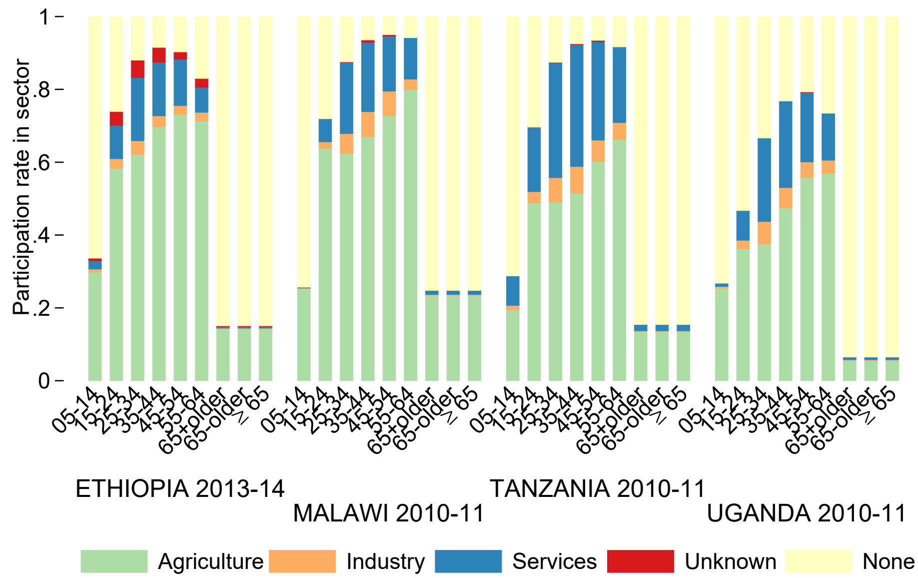


Figure 2.10: Primary sector participation rates across age cohorts.

terms of person-days, than do men. The average years of education completed tend to be highest for services sector workers and lowest for agricultural sector workers. These educational differences point to possible systematic cross-sector skills differences. Individuals who do not supply labor to any sector are younger and more female, on average, than agricultural workers.

Figure 2.10 depicts the changing primary sector of workers across all major age cohorts. Youth (ages 15-24) have lower participation in economic activities than do young adults (ages 25-34). Economically active youth supply larger shares of labor to agriculture (compared to industry and services) than do economically active young adults. Despite these differences, labor shares are robust to the specification of the adulthood threshold at age 25 rather than age 15.[10]

Tables 2.4 and 2.5 summarizes individuals participation in self and wage employment activities by sector, describing participation rates and basic char-

[10]Results can be shared upon request.

acteristics of participants in both rural (Table 2.4) and urban (Table 2.5) populations. The tables summarize all individuals who participate in self and wage employment in each sector, not just those who primarily participate. Davis, Di Giuseppe, and Zezza (2014) generate household level estimates of participation in wage and self-employment by rural LSMS-ISA households.

Table 2.4: Characteristics of rural own account and wage workers

	ETHIOPIA 2013-14		MALAWI 2010-11		TANZANIA 2010-11		UGANDA 2010-11	
	Self	Wage	Self	Wage	Self	Wage	Self	Wage
Agriculture (share)	0.863	0.0588	0.886	0.326	0.817	0.0897	0.741	0.152
Hours / year, mean	553.3	631.5	423.1	432.2	561.5	382.6	638.3	722.9
Share female	0.516	0.39	0.527	0.439	0.53	0.379	0.568	0.491
Age, mean	40.03	39.66	39.17	37.21	41.15	39.3	40.65	39.8
Educ yrs, mean	1.51	1.614	4.899	4.589	5.502	5.13	5.416	4.633
Returns / year (positive), med	439.1	168.3	367.1	92.28	330.4	164	246.2	72.42
Returns / year (positive), mean	552.9	427.8	422.4	265.4	425	513.2	322.3	408.2
Industry (share)	0.027	0.008	0.298	0.034	0.051	0.027	0.033	0.064
Hours / year, mean	1129	1339	586.2	1371	490.5	1138	581.5	912.9
Share female	0.657	0.18	0.529	0.102	0.521	0.176	0.652	0.34
Age, mean	37.57	37.1	40.34	37.26	40.39	38.05	40.81	38.22
Educ yrs, mean	2.085	3.8	4.872	6.852	6.883	7.656	5.156	6.625
Returns / year (positive), med	260	1033	199.9	769	394.3	888.1	253.5	348.5
Returns / year (positive), mean	617.4	1782	423.9	1475	896.6	3189	690.9	2063
Services (share)	0.083	0.036	0.353	0.064	0.256	0.086	0.132	0.118
Hours / year, mean	1336	1480	926.7	1319	647.3	1433	1448	1240
Share female	0.559	0.253	0.399	0.207	0.526	0.269	0.418	0.359
Age, mean	37.51	35.97	36.36	38.86	39.23	38.9	38.88	39.64
Educ yrs, mean	2.601	8.975	6.484	9.811	6.77	9.214	7.397	9.455
Returns / year (positive), med	473.5	1946	291.6	1102	546.5	2049	633.7	937.7
Returns / year (positive), mean	1245	2785	674.7	1982	1737	4033	1562	2334

Table 2.5: Characteristics of urban own account and wage workers

	ETHIOPIA 2013-14		MALAWI 2010-11		TANZANIA 2010-11		UGANDA 2010-11	
	Self	Wage	Self	Wage	Self	Wage	Self	Wage
Agriculture (share)	0.0482	0.0154	0.284	0.204	0.257	0.0198	0.236	0.0632
Hours / year, mean	307.7	1641	183.5	860.6	354.9	1016	376.6	503.6
Share female	0.45	0.366	0.525	0.264	0.562	0.321	0.607	0.54
Age, mean	40.14	39.27	37.57	37.05	41.67	36.64	42.1	40.67
Educ yrs, mean	5.455	9.244	8.231	7.08	7.097	5.786	8.325	8.323
Returns / year (positive), med	336.3	2069	240.1	430.7	259.4	512.4	141.1	154.5
Returns / year (positive), mean	524.9	2833	303.8	1055	337.2	1468	279.9	457.7
Industry (share)	0.028	0.0931	0.104	0.0807	0.0842	0.0705	0.0462	0.0977
Hours / year, mean	1512	1880	1256	1814	557.5	1653	1530	1319
Share female	0.597	0.25	0.387	0.0787	0.424	0.15	0.451	0.235
Age, mean	39.78	36.49	36.81	37.1	38.24	36.88	40.75	37.13
Educ yrs, mean	6.627	8.575	9.387	10.73	7.638	8.942	9.149	9.797
Returns / year (positive), med	903.9	2367	1268	1964	755.1	2186	461.3	1432
Returns / year (positive), mean	3459	3717	2636	4559	1453	5820	1153	3914
Services (share)	0.195	0.316	0.516	0.26	0.397	0.245	0.238	0.242
Hours / year, mean	1692	1888	1232	1846	730.6	2084	2057	1845
Share female	0.494	0.417	0.472	0.286	0.583	0.306	0.576	0.425
Age, mean	37.59	36.02	34.71	36.73	38.61	37.32	40.54	38.26
Educ yrs, mean	6.951	10.92	9.349	11.36	8.158	9.765	9.261	11.91
Returns / year (positive), med	1549	2540	879.3	2076	1168	3074	1542	3331
Returns / year (positive), mean	3542	3990	3665	5891	2928	5665	2918	5814

Participation in self-employment is more common than wage labor participation in all countries, with 74-89% of rural adults participating in farming. Agricultural wage labor participation is less common than farming, with fewer than 15% of rural adults participating in Ethiopia, Tanzania and Uganda, and 33% in Malawi. In all countries, the average agricultural wage laborer is much more likely to be male than the average farm worker. Agricultural wage workers have more education, on average, than farm workers in Ethiopia and Malawi, and less in Tanzania and Uganda.

Behind agriculture, the services sector has the next highest overall participation rate. And in urban areas, the services sector is the most important. Workers are more likely to participate in the services sector through self-employment than wage employment in both rural and urban areas, except in urban Ethiopia where wage employment is higher than self-employment. Wage labor participants in the services sector are more likely to be male and to have higher education levels than self-employed service sector participants. Both wage and services sector participants supply similar numbers of hours per year except in Tanzania, where service sector self-employed workers supply far fewer hours than do wage laborers.

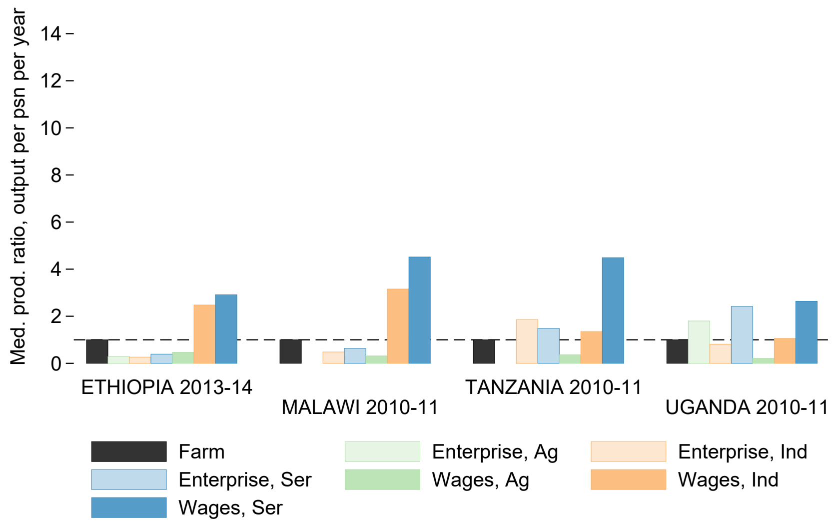
Within the industry sector, rural individuals typically are more likely to participate as self-employed rather than wage workers. Participation as either self-employed or wage workers is below 6.5% everywhere except Malawi, where one third of rural adults participate in industry sector self-employment.[11] In rural areas, industry wage laborers are very strongly male, while industry self-

[11] In the Malawi survey, industry-sector self-employment firms tend to be involved in rudimentary value addition activities like beer brewing or brick making or basket weaving. In Malawi, basic roadside selling of food products was coded as manufacturing, whereas in other countries, these activities are usually coded to the service sector.

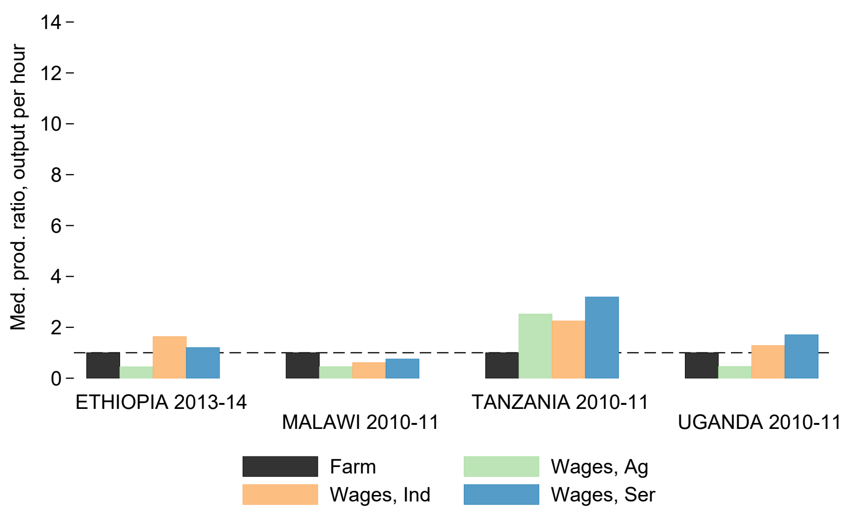
employed workers are mostly female. Urban participation in industry sector activities is slightly higher than is rural participation everywhere except Malawi.

There is always concern that differences in productivity may reflect observed and unobserved differences in the households and individuals participating in the activities rather than inherent differences in the economic productivity of the activities themselves. In an attempt to control for the effects of household level selection on productivity gap measurement, I generate within-household sector-activity productivity gap measures for households that participate in multiple activities. For households that participate both in farming and another activity, these reflect the ratio within the household between returns to the non-farm activity and farming. Figure 2.11 depicts the median of intra-household productivity gaps. The conditional productivity gap for service sector wages, for example, is based on a comparison between annual returns per participant to farming and service sector wage labor within households who participate in both activities.

The conditional sector-activity gaps depicted in Figure 2.9 are considerably smaller than the unconditional gaps depicted in Figure 2.7. The fact that within-household gaps are so much smaller than between-household gaps suggests that heterogeneity in household characteristics between activity-sector participants could partly explain between-household productivity gaps. The observed small magnitude of intra-household gaps also suggests that structural barriers to improved household productivity that span across sectors may constrain households opportunities to raise their productivity levels. Such structural barriers to improved household productivity, including but not limited to the difficulties of accumulating human capital, would limit the opportunity to achieve



(a)



(b)

Figure 2.11: Conditional productivity gaps by activity for all households (median). Panel (a) depicts the median of intra-household productivity ratios between farming and other activities, where productivity is defined as output per worker per year. Panel (b) depicts the median of per hour intra-household productivity ratios. This analysis is based only on households that participate in farming and another activity.

productivity growth by shifting labor out of agriculture, even though productivity appears higher outside of agriculture on a per capita basis.

2.6.2 Farm and Non-Farm Linkages

Tables 2.6 and 2.7 break down non-farm self and wage employment activities into a more granular list of sectors. For the self-employment columns, the total number of households in the dataset is provided, along with the number of households that operate at least one non-farm enterprise, and the total number of firms present in the dataset. This final number is larger because some households operate more than one firm. These firms are then categorized by ten sub-sectors of the economy. In the non-farm wage employment columns, the total number of individuals of working age is listed, along with the total number who participate in wage labor, and the total number of jobs reported in the dataset. Again, because some individuals have more than one source of wage-earning income, the number of wage earning jobs is larger than the number of wage market participants. Industry sector activities are divided into mining, manufacturing, electricity and utilities, and construction. Service sector activities are broken into commerce, transport and communication, general services, and finance. Summaries of activities are provided separately for rural and urban areas (Tables 2.6 and 2.7, respectively). Many of the activities do not occur, or occur only once, in each sample.

Table 2.6: Detailed sectors of non-farm self and wage employment, Rural areas

	ETHIOPIA 2013-14		MALAWI 2010-11		TANZANIA 2010-11		UGANDA 2010-11	
	Self	Wage	Self	Wage	Self	Wage	Self	Wage
N (HHs or indivs) in sample								
Households	3,776		2,390		2,583		2,049	
Individuals		5,941		3,428		4,331		3,451
of which N participate	1,263	1,746	441	1,354	1,044	786	954	976
N firms (self) or jobs (wage)	1,683	1,956	469	1,415	1,404	856	1,473	1,102
Share of firms (self) or jobs (wage) by sub-sector								
Ag and Primary Prod	0.0778	0.2240	0.0085	0.7760	0.0135	0.4050	0.0978	0.5340
Mining	0.0297	0.0041	0.0107	0.0035	0.0135	0.0152	0.0129	0.0036
Manufacturing	0.1600	0.0067	0.3820	0.0269	0.1410	0.0350	0.1320	0.0853
Electricity, Utilities		0.0041		0.0021		0.0117	0.0014	0.0018
Construction	0.0119	0.0174	0.0085	0.0297	0.0071	0.0467	0.0020	0.0617
Commerce	0.6350	0.0046	0.4990	0.0155	0.6620	0.1160	0.2630	0.0799
Transport, Storage, Comm.		0.0072	0.0277	0.0064	0.0256	0.0397	0.0299	0.0299
Finance, Real Estate	0.0018	0.0067		0.0021	0.0014	0.0187	0.0007	0.0000
Services	0.0547	0.1190	0.0640	0.1330	0.1310	0.3080	0.1150	0.2030
Other Industries		0.0092		0.0000		0.0000		0.0000
Missing sector info	0.0285	0.5970		0.0050	0.0043	0.0035	0.3460	0.0000

Table 2.7: Detailed sectors of non-farm self and wage employment, Urban areas

	ETHIOPIA 2013-14		MALAWI 2010-11		TANZANIA 2010-11		UGANDA 2010-11	
	Self	Wage	Self	Wage	Self	Wage	Self	Wage
N (HHs or indivs) in sample								
Households	3,776		2,390		2,583		2,049	
Individuals		5,941		3,428		4,331		3,451
of which N participate	1,263	1,746	441	1,354	1,044	786	954	976
N firms (self) or jobs (wage)	1,683	1,956	469	1,415	1,404	856	1,473	1,102
Share of firms (self) or jobs (wage) by sub-sector								
Ag and Primary Prod	0.0778	0.2240	0.0085	0.7760	0.0135	0.4050	0.0978	0.5340
Mining	0.0297	0.0041	0.0107	0.0035	0.0135	0.0152	0.0129	0.0036
Manufacturing	0.1600	0.0067	0.3820	0.0269	0.1410	0.0350	0.1320	0.0853
Electricity, Utilities		0.0041		0.0021		0.0117	0.0014	0.0018
Construction	0.0119	0.0174	0.0085	0.0297	0.0071	0.0467	0.0020	0.0617
Commerce	0.6350	0.0046	0.4990	0.0155	0.6620	0.1160	0.2630	0.0799
Transport, Storage, Comm.		0.0072	0.0277	0.0064	0.0256	0.0397	0.0299	0.0299
Finance, Real Estate	0.0018	0.0067		0.0021	0.0014	0.0187	0.0007	0.0000
Services	0.0547	0.1190	0.0640	0.1330	0.1310	0.3080	0.1150	0.2030
Other Industries		0.0092		0.0000		0.0000		0.0000
Missing sector info	0.0285	0.5970		0.0050	0.0043	0.0035	0.3460	0.0000

Next, I use respondents free descriptions of their self-employment and wage employment activities, along with the detailed industry codes assigned by enumerators, to examine carefully the kinds of non-farm activities in which respondents are involved. Within each sub-sector, I use text analysis to identify the words that most commonly appear in respondents descriptions (Laver, Benoit, and Garry, 2003). We do not observe the level of formality associated with household firms and wage-earning jobs because there is not enough comparability across survey questionnaires to describe formality of employment arrangements and/or firm registration.

In rural areas of all four countries, agricultural wage labor is the largest category of wage employment. In Ethiopia, around 60% of wage employment occurs as casual or informal labor for which no sector information or job description was collected. Most likely, this labor is supplied to the agriculture sector.[12] Based on text analysis of the descriptions provided, most agricultural jobs involve casual labor on farms for food or cash crop production, or they involve livestock tending, hunting, fishing, and collection of forestry products, such as fuel wood. Agricultural sector non-farm self-employment, which is not common, also tends to involve production of livestock, fishery, or forestry products.

Within the industry sector, mining does not play an important role in rural or urban areas of any of the datasets we analyze. Manufacturing accounts for between 13% and 38% of rural NFEs, with the smallest share in Tanzania and the largest in Malawi. However, only 1-9% of wage-earning jobs occur in manufacturing. According to text analysis, manufacturing NFEs focus heavily on

[12]This impression is based on my own experience shadowing LSMS-ISA survey teams in Ethiopia during the 2012-2013 round of surveys, and on discussions with survey teams in Ethiopia. Ganyu wage labor from Malawi and Productive Safety Net Program labor from Ethiopia were categorized as agricultural based on consultations with survey teams.

elementary activities such as brewing alcoholic beverages, charcoal production, milling grains, butchering and other agricultural processing, baking and other value addition activities, and the production of household goods, clothing, and other handicrafts. Manufacturing wage jobs are similar, with a focus on agri-processing for cash crops, timber, and textiles, as well as the manufacturing of bricks and other building materials. Construction accounts for between 2% and 6% of rural wage jobs and between 5% and 9% of urban wage jobs but fewer than 2% of NFEs. Construction wage employment, according to text analysis, typically involves working as a laborer on a building or road construction site.

Individuals and households who participate in the industry sector are involved mainly in manufacturing activities that have strong links with primary agricultural production. Industry sector participants contribute to manufacturing raw agricultural materials into typically non-tradable goods meant for local consumption. These patterns suggest strong links between rural industry-sector activities and agriculture. In rural areas, the manufacturing industry stands to gain from productivity growth in agriculture, and rural manufacturing workers are poised to benefit from demand spurred by rising agricultural incomes in rural areas. Because the manufacturing activities reported in these surveys are so closely linked with agriculture, one would not expect to see expansion of rural industry sector activities independently, without any agricultural growth. These classic Mellor-Johnston linkages are quite prominently featured in rural households economic activities.

Commerce is the dominant focus of self-employment in the services sector, while jobs tend to involve general services provision. Commerce comprises between 26% and 66% of both rural and urban firms. These are involved in activities such as the wholesale and retail trade of fruits and vegetables, other food items, charcoal, second hand goods, and other household goods. Commerce accounts for up to 20% of wage earning in urban areas of Tanzania, but the share is more often closer to 5%-10% in urban areas, and lower in rural areas. Commerce wage earners are most commonly sales clerks and store attendants. The general services category is the most important for wage employment, accounting for 42%-45% of urban jobs across all countries, and 12%-31% of rural wage earning jobs. These wage workers include teachers, health, social and religious workers, public administrators, technicians, domestic service providers, as well as restaurant, hotel, and tourism employees. General services account for a smaller share of firms than of wage jobs. The most common firm descriptions include restaurants, caterers, bars, hotels, professional service providers, and repair shops. The transport sub-sector accounts for a small share of self and wage employment everywhere. Transport activities tend to focus on transportation services provided by bicycle, taxi, bus, or vehicle. Finance and real estate are almost nonexistent in rural areas and account for 1%-3% of urban wage earning jobs, which are most commonly administrative in nature.

Buying and selling agricultural products comprises a large share of commerce activity, with respect to both self and wage employment. As with the industry sector, the services sector activities in which rural households participate are non-tradable in nature, and very focused on local consumers. Because these service sector activities serve local consumers whose incomes are dominated by agriculture-sector activities, the Mellor-Johnston linkages are again quite prominent. One would expect agricultural productivity growth to spur demand for increased local service sector labor. Given the nature of service sector activities, it would be hard to imagine strong growth in the services sector absent agricultural growth.

2.7 Conclusion

Micro level cross-sector labor productivity gaps are smaller than those generated using national accounts data. Inter-sectoral differences in annual earnings per worker arise from differences in employment volume (hours per worker of labor supplied) rather than wages or productivity per hour of labor supplied. At least half of these per-worker productivity gaps can be explained by differences in hours worked across sectors. The tendency is for individuals participating in agriculture to supply fewer hours to agriculture, on average, than individuals participating in other sectors. Returns to an hour of labor supplied outside of agriculture are about 1.4 times as high as returns to an hour of agricultural labor, on average, in the four countries analyzed.

Generally, the micro evidence seems consistent with the idea that there is some scope for achieving productivity gains by shifting labor from agriculture to industry or services. Households expect industry and service sector wage workers to earn higher returns per year than farm workers. Self-employment brings higher annual returns to participants than farming but lower than wage employment. Since micro gaps are smaller than macro gaps, workers, who are the owners of labor, may not stand to reap the large benefits of labor exiting agriculture that are expected in the economy as a whole (should national accounts data indeed reflect true economy-wide productivity gaps). Small per-worker-per-year micro gaps also suggest that agriculture-sector workers do not feel as strong a pull from industry and services as one might expect based on national accounts data. Small per-hour gaps do not undermine agricultures role in structural transformation. Despite low per-hour gaps in agriculture, it appears that workers have an excess of labor that could be absorbed productively in other sectors. This requires growth in demand for labor within or outside of agriculture.

Though underemployment in agriculture has been observed in the developing country context, it is not a well understood phenomenon. Widespread underemployment could erode the benefits of using agricultural labor more productively. The existence of large employment gaps across sectors raises the question of what limits the supply of hours in agriculture and what role technology, infrastructure and policies might play in addressing agricultural underemployment. Smallholders could be operating at high levels of technical efficiency, yet face environmental production constraints, such as limiting in-season rainfall for rainfed crops (Sherlund, Barrett, and Adesina, 2002; Schultz, 1964). In this case, there could be scope to smooth labor demand with water control infrastructure and management practices. Demand for agricultural labor could be constrained due to the time-sensitive nature of agricultural tasks, such as land preparation, planting, weeding, and harvest. If certain time-sensitive tasks create labor supply bottlenecks, then interventions to address these bottlenecks, such as mechanization, could generate productivity gains. Mechanization has been very limited in LSMS-ISA countries (Sheahan and Barrett, 2014), though this demand could reflect frictions in capital markets.

Barriers to participation in non-farm self or wage employment could limit labor supply outside of agriculture by underemployed agricultural workers (Barrett, Reardon, and Webb, 2001; Rodrik, 2014b). These could arise from constraints to accumulating human capital, or limited opportunities for off-farm employment. The evidence suggests that individuals and households may indeed face barriers to participation in non-agriculture activities. Workers who primarily participate in the industry and service sectors tend to also participate in agriculture, while the reverse is not true of workers who are primarily agricultural. Service sector participants, in particular, tend to have higher education levels than workers in other sectors. Some households may face structural barriers to labor productivity growth that span across multiple sectors. The small size of conditional gaps (within-household gaps faced by participants in multiple sectors) relative to unconditional gaps (pooled, cross-sector gaps) suggests that selection effects into non-agriculture activities contribute to cross-sector productivity differentials. Households who are unable to diversify might face even smaller productivity gains outside of agriculture than those who are, further eroding the benefits of structural reallocation of labor.

Overall, the analysis emphasizes agricultures key role in Sub-Saharan African economies, while also raising questions about agricultural employment gaps, their determinants, and how they shape the opportunity to achieve economy-wide labor productivity growth. A between-sector gradient in annual output per worker remains to be exploited. Improving annual output per worker within agriculture, the highest participation sector by far, requires a better understanding of labor demand by smallholder farmers.

Agriculture, and specifically the operation of household farms, remains a dominant economic activity in rural areas of Sub-Saharan Africa. And, furthermore, much of the labor supplied to industry and services sectors involves the processing and trading of agricultural and other primary goods for consumers whose incomes are dominated by agriculture-sector activities. Furthermore, the non-farm activities in which rural households are involved, across countries, are incredibly closely linked with agriculture. These strong links highlight additional benefits to achieving agricultural productivity growth, which can increase the supply of raw materials for manufacturing and increase the demand for non-tradable goods and services. These linkages are also sobering as, apart from agriculture, no engine for rural economic growth is apparent.

CHAPTER 3

OCCUPATIONAL CHOICE AND AGRICULTURAL LABOR EXITS IN SUB-SAHARAN AFRICA

3.1 Introduction

Economic development is characterized, almost universally, by rising output per agricultural worker and the movement of labor from agriculture to other sectors, which together result in rising incomes and falling incidence of poverty (Timmer, 2009). African countries are mostly in the early stages of this structural transformation process, with large cross-sector productivity gaps and large labor shares still in agriculture (Gollin, Lagakos, and Waugh, 2014b). Recently, though, growth has been observed in annual output per worker across Sub-Saharan Africa. In the aggregate, labor exits from agriculture to other sectors explain about half of the observed increases in annual output per worker (McMillan and Harttgen, 2014).

Growth in agricultural labor productivity is closely associated with poverty reduction, both through the direct effects on the many workers who participate in the agricultural sector, and indirectly, because it leads to growth in non-agriculture sectors and lowers food prices through increased per capita food supplies (De Janvry and Sadoulet, 2010). Few debate the importance of farming to poor households, simply because farming is the occupation in which the poor participate with the highest frequency (Christiaensen, Demery, and Kuhl, 2011).

What is under debate in Sub-Saharan Africa today, however, is the scope for achieving structural change through smallholder-focused interventions in

Sub-Saharan Africa. Some agriculture-skeptics argue that smallholder farmers are weak agents for labor productivity growth of the magnitude necessary to trigger large scale poverty reduction due to low baseline productivity and poor prospects for improving labor productivity within agriculture (e.g., Dercon, 2013; Collier and Dercon, 2014; Dercon and Gollin, 2014). By extension, these skeptics question the role of agricultural interventions in poverty reduction.

Historically, technology-led agricultural productivity growth has been the essential lever for launching structural transformation (e.g., Johnston and Mellor, 1961; World Bank, 2008; Christiaensen, Demery, and Kuhl, 2011). The economy-wide labor productivity growth that accompanied the widespread adoption of high-yielding varieties in South and East Asia and Latin America during the Green Revolution serves as evidence (Evenson and Gollin, 2003). And most economists have long rejected the idea that economic growth can be spurred in poor economies while agriculture remains stagnant (Ranis, 2004). Nevertheless, development experts have highlighted the importance of interventions that raise labor productivity more generally and in other sectors of the economy.

Today's debate on agriculture's role in overall economic growth in Sub-Saharan Africa hinges on the potential for raising labor productivity in agriculture and in other sectors, and on the impacts of rising labor productivity. In this case, labor productivity in a sector refers to the net returns achieved per worker who participates in that sector. Many types of interventions are associated with rising labor productivity – agricultural technology, improved education, improved infrastructure, etc. The effects of labor productivity enhancing interventions can play out on the intensive margins, for workers who remain in the same occupation as productivity changes, and on the extensive margins, as workers shift occupations in response to productivity changes.

The occupational choice decisions that underlie the structural transformation process play out among many households and farms in heterogeneous settings. While there is empirical regularity in the aggregate relationships between agricultural productivity growth, non-farm growth, agricultural labor exits, overall economic growth, and poverty reduction; the micro-economic processes that underlie these relationships are not well understood (Foster and Rosenzweig, 2007). To my knowledge, no empirical study has explicitly examined the micro-economic dimensions of agricultural transformation in Sub-Saharan Africa in the context of occupational choice and technological change.

One major reason for research scarcity on this topic has been, until very recently, lack of datasets that cover relevant farming and non-farming activities of households in both urban and rural areas, including household-managed non-farm enterprises and farms as well as wage labor. Taking advantage of newly available, innovative Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) datasets, I examine the role that improved agricultural technology plays in fostering structural change in African economies. I match LSMS-ISA datasets with a number of other relevant datasets using geo-referenced household locations. I then model annual household returns to participation in farming, non-farm self employment, and wage employment. I find that, in farming, latent labor productivity for households is closely related to household size, the cost of hired labor, land owned, precipitation, and soil nutrients. Self employment latent labor productivity is closely related to peri-urban status, the age of the household head, and ownership of productive assets. And wage employment latent labor productivity is closely related to market access, male headedness of the household, education wtihin the household, and local wage rates.

I use imputed latent labor productivity measures to estimate a household level polytomous model of occupational choice. Predicted occupational choices closely match actual occupational choice shares for observations left out of the estimation sample, and for different sub-populations within the estimation sample. Finally, I simulate the welfare impacts of doubling labor productivity in farming, self employment and wage employment, respectively. I estimate these impacts both along the intensive margins of participation, for households that do not change occupational choices, and along the extensive margins of participation, for households that do change occupational choices. The lion's share of welfare effects are experienced by households that do not shift occupational choice. Participation in farming is overall non-responsive to any of the productivity shocks. Some entry into self employment is seen for the self employment labor productivity shock, and into wage employment for the wage labor productivity shock. Households tend to enter into self and wage employment without exiting farming. The results suggest that agricultural labor productivity growth can lead to large welfare gains because so many households participate in farming, without impacting the probability that households participate in farming.

3.2 Model

In Sub-Saharan Africa, workers outside of agriculture tend to have higher returns per worker per year (McCullough, 2015; Gollin, Lagakos, and Waugh, 2014b; McMillan and Harttgen, 2014). This occurs not because activities outside of agriculture are inherently more productive per hour of labor worked, but because workers outside of agriculture tend to supply more hours of labor per year, while the agricultural sector houses a large reservoir of underemployed workers (McCullough, 2015).

Because ability to participate in fuller-employment activities outside of agriculture seems to be a very important determinant of annual worker returns and household expenditures per capita, this paper focuses on the extensive margin of labor supply (choice of occupation) rather than the intensive margin (hours worked per year in each occupation). I use a discrete choice framework not only because the extensive margins of labor supply are of greater interest than the intensive margins in the structural change framework, but also because labor supply is difficult to measure on the intensive margin, with measurement error differing systematically across occupational choices.

While self employment is not as commonly included in occupational choice models as is wage employment, I allow for it as an occupational choice because it is a very common one in the Sub-Saharan African setting. Furthermore, the shift of labor from self employment to wage employment is a key characteristic of the development process (Behrman, 1999), and one that is associated with labor productivity growth even when workers do not change sectors as they shift from self to wage employment (McCaig and Pavcnik, 2013). Here, self employment does not include own production of household goods, such as child rearing, but rather the operation of household-managed enterprises intended to generate income for the household.[1] I also allow households to participate in multiple activities simultaneously, reflecting the reality of occupational choices observed in this setting (Barrett, Reardon, and Webb, 2001; Davis et al., 2010).

I assume a representative household makes its occupational choice of participation ($P = 1$) or non-participation ($P = 0$) in each of three activities: farm operation (F), wage employment (WE), and self employment (SE). Allowing for each binary option in that triplet, the choice set contains $8 = 2^3$ possibilities. Households derive utility from income per household member (income per adult equivalent is Y_i, with s_i denoting household size).

I use a basic household random utility model with discrete occupational choices to derive estimable equations for structural model parameters, where household utility has both observed and unobserved components, and households are assumed to select the option that brings it the highest utility. Random utility occupational choice models are widely used in labor economics to study the effects of policies and taxes on labor supply (e.g., Keane and Wolpin, 1997; Keane and Moffitt, 1998; van Soest, Das, and Gong, 2002). I model occupational choices at the household level rather than the individual level because, for two of the three activities available to households – farming and self employment – returns are only observable at the household level.[2] In this formulation, utility received by household i from decision j (u_{ij}) is known to the decision-maker

[1]Virtually all households have at least one member who engages in the production of household goods on the extensive margin, so participation in household good production would not be very interesting to model empirically, at least at the household level.

[2]Modeling individual occupational choices would be an interesting extension, which would allow for closer examination of age and gender patterns. It requires use of some assumptions on how returns to participation in farming and self employment vary within the household.

(household i) but not to the researcher. Household i chooses option k if and only if $u_{ik} \geq u_{ij}\ \forall\ k \neq j$ and $u_{ik} > u_{ij}$ for at least one $k \neq j$. The household observes its own utility (u_{ij}) across choices, which can be decomposed into a component observed by the researcher (U_{ij}) and an unobserved compoenent (ε_{ij}). Some distributional assumptions made on ε_{ij} are required for maximum likelihood estimation of model parameters. The assumptions made here are discussed in section **??**.

The household's decision follows:

$$\max_{P_{ij}=(P^F,P^{WE},P^{SE})} \quad u_i(P_{ij}) = \alpha \cdot Y_{ij}^{.5} + \gamma_j \cdot C_i + \delta_j + \varepsilon_{ij} \tag{3.1}$$

s.t.

$$Y_i \equiv \frac{1}{s_i}(\Pi_i + R_i) \tag{3.2}$$

$$\Pi_i \quad \equiv \quad P_i^F \cdot \Pi_i^F \quad + P_i^{SE} \cdot \Pi_i^{SE} \quad + \quad P_i^{WE} \cdot \Pi_i^{WE} \tag{3.3}$$

$$P_i^a \in \{0, 1\} \qquad \forall \quad a \in \{F, WE, SE\} \tag{3.4}$$

Each household's income (Y_i) is determined by the process defined in equation 3.2. For each household and occupational choice, the corresponding income is determined by the net returns (profits) to participating in farming, wage employment, and self employment (Π^F, Π^{WE} , and Π^{SE}, respectively). Income also includes non-labor income sources (R), which do not vary across occupational choices and are derived from public and private transfers and other sources. The index variable j refers to each of the 8 unique combinations of participation in the three different activities from which the household selects its occupational choice. The household's choice of occupation is influenced by additional choice predictors (C_i), such as the household head's parent's education level, that influence a household's selection into an occupation apart from affecting the returns to participation.

There is no leisure consumption in the model. Rather, any dis-utility associated with supplying labor to occupation is reflected in the occupation-specific preference shifters (δ_j). I use a functional form that is monotonically increasing and concave in income.[3] I do not impose *a priori* that utility is decreasing in labor supply. In this model, both willful non-participation in the labor force and unemployment (unsuccessfully attempting to participate in wage employment or other activities) are observed equally, as non-supply of labor. It is not possible to distinguish between these outcomes empirically.

Net returns to participation in an activity are determined by a flexibly specified indirect profit function, which is the dual of a multi-input production function. Consider a total of K farm inputs and outputs, hereafter the netput vector. The profit function takes the input and output price vector as arguments. Here, I use a flexible Generalized Leontief form to specify the returns to activity $a \in \{F, SE, WE\}$. This functional form is advantageous for its flexibility. This process is described in equation 3.5.

$$\Pi_i^a(P_i^a) = \sum_{k=1}^{K} \beta_k^a x_k^{1/2} + \sum_{k=1}^{K} \sum_{m=1}^{M} \beta_{km}^a (x_k^{1/2} x_m^{1/2}) + e_{ai} \tag{3.5}$$

Here, x_k refers to the k^{th} variable in the netput vector, which includes the variables that proxy for household shadow prices and relevant context variables that condition household returns to participation in an activity. For example,

[3]The utility function parameters are quite robust to different specifications that do not impose that utility is concave or monotonically increasing in income.

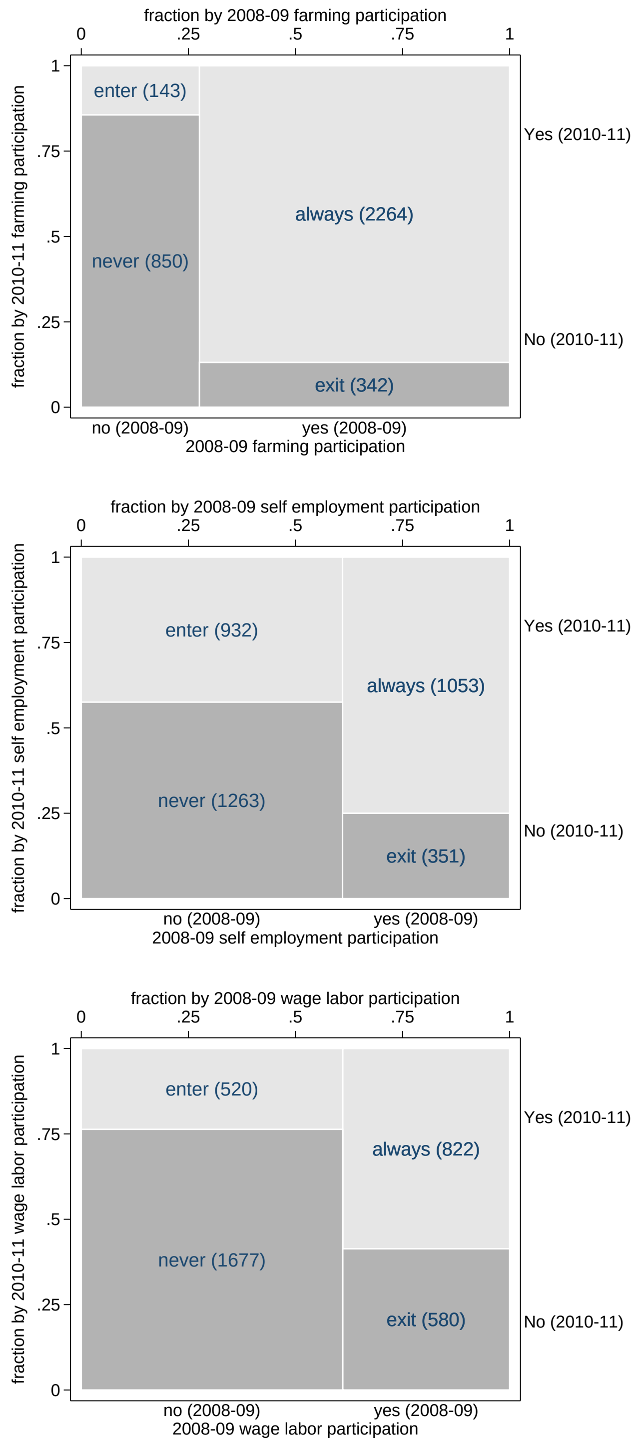
mean rainfall is used as a control for returns to farming. All of the variables are interacted with each other in the specification. I model returns to sector participation using a stylized profit function rather than an expenditure function or production function for several reasons. First, it allows me to avoid modeling endogenous input use decisions, which then lead to an infinite choice set of inputs used and occupational choices. Rather, the stylized profit function takes prices and key context variables as arguments, which are observable for households that participate in an activity and for those that do not. Second, this approach is relevant in the developing country setting, where there is strong evidence of input and output market frictions, and different households face different prices (Dillon and Barrett, 2014; De Janvry, Fafchamps, and Sadoulet, 1991; Barrett, 2007). Rather than restricting the choice set of occupations available to different households, I allow shadow prices and returns to vary as a function of geographically determined and household specific observable variables. The specific instruments included for shadow prices are discussed in section **??**.

Neither the profit function nor the occupational choice model explicitly includes the fixed costs associated with entering or exiting an occupation. These costs are simply not available in the data. When fixed costs associated with participating in an activity are ignored, discrete choice labor supply models tend to under-predict non-participation in that activity (van Soest, Das, and Gong, 2002). Occupation-specific preference shifters (δ_j) can pick up these fixed costs when they are not observed directly. The consequence is that one cannot then disentangle the effects of fixed costs versus alternative-specific preference heterogeneity on participation choices. This must be considered when interpreting the vector of preference shifting parameters.

This model rests on several assumptions. It is a static model, and therefore does not allow for borrowing or saving. Risk and uncertainty are not featured in the profit functions for farm and non-farm enterprises, though risks that households associate with different occupations, and preferences for those risks, can be absorbed into the occupation-specific preference shifters. At this point, general equilibrium effects on wages and prices are not explored. Relevant equilibrium effects in the structural change context include employment effects resulting in changing wage rates, changes in relative output prices due to non-homothetic demand and non-tradability of goods and services consumed (or closed markets). These equilibrium effects are certainly of interest in future studies. The partial equilibrium estimates remain interesting and relevant in the short term.

I estimate a static, rather than dynamic, model because the time interval between survey rounds is fairly short (2-3 years). Transition matrices between the first and second survey rounds for farm, self and wage employment are shown in Figure 3.1. Overall, a plurality of households never participate in wage labor markets and more households appear to exit wage labor employment than enter it. Conversely, more households enter self employment than exit it. The largest categories are households who do not change participation in any given activity across survey rounds. This is particularly true for farming, where a large majority (about two thirds) of households farmed in both survey rounds. Furthermore, there is not a lot of temporal variation in many of the variables used to estimate returns to participation. The focus is on explaining, in a pooled cross-section, observed patterns of occupational choice within the framework of structural change processes, and then to address how these patterns might change in different circumstances.

Figure 3.1: Transition matrices for households between 2008 and 2011, by activity. Farming is shown, followed by self employment and wage employment.



3.3 Estimation

Estimation proceeds in two stages. In the first stage, I estimate profit function parameters. In the second, I estimate the parameters of an occupational choice model using imputed profits. One major challenge in estimating returns to participation is, of course, that returns are only observed for households that elect to participate. I control for selection effects by estimating returns to participation on the full sample of participants and non-participants, using a Heckman selection model (Heckman, 1979). For each activity (farming, wage employment, and self employment), I estimate annual returns per household as a function of the x variables described in equation 3.5 and the selection variables (C) described in equation 3.1. The estimation equation follows. Equations 3.6 and 3.7 are estimated jointly, and u_1 and u_2 are have a correlation coefficient of ρ.

$$\Pi_i^a(P_i^a) = \sum_{k=1}^{K} \beta_k^a x_k^{1/2} + \sum_{k=1}^{K} \sum_{m=1}^{M} \beta_{km}^a (x_k^{1/2} x_m^{1/2}) + u_1 \tag{3.6}$$

$$\text{and } \Pi^a \text{ is observed if:} \lambda_j \cdot C_i + u_2 > 0 \tag{3.7}$$

Then, using the estimated β parameters, I impute returns to participation in farming, wage, and self employment for all households, regardless of their participation. Imputed returns are then used to generate for each household a vector of incomes, one for each of the 8 possible choices. I assume that non-participation in activity a results in a profit of 0 for that activity.

For the second stage estimation, I use a mixed logit model in order to avoid the strong independence from irrelevant assumptions that occur with multinomial logit models. By estimating the preference shifters as random coefficients, I allow for preference heterogeneity and correlation of errors across choices. The random coefficient is δ_{kj}, and it is estimated at the lowest administrative level above the household. Because there is only one observation per household in the sample, it is not tractable to estimate the random coefficient at the household level. This approach is akin to an error components model, with δ_{kj} serving as a structured component of the unobserved utility (Train, 2002). The remaining component of the unobserved utility, error term ε_{ij}, is assumed to be independent and identically distributed according to the extreme value (Gumbel) distribution.

After integrating out the random error, the probability of each choice is then given by equation 3.8. The index term w refers to ward level, which is the lowest administrative level observed in the data. Since there is only one observation per household, I estimate the random parameters at the ward level rather than the household level. Because of the considerable computational demands of exact maximum likelihood estimation, I use maximum simulated likelihood to estimate α, γ, $\bar{\delta}$, and Σ (Gu, Hole, and Knox, 2013).

$$\text{Prob}(P_{ijw} = 1) = \int \Big(\frac{e^{\alpha Y_{ij}^{.5} + \gamma_j' c_i} \cdot e^{\delta_{jw}}}{\sum_k e^{\alpha Y_{ik}^{.5} + \gamma_k' c_i} \cdot e^{\delta_{kw}}}\Big)\phi(\delta)d\delta \tag{3.8}$$

$$\delta \sim \mathcal{N}(\bar{\delta}, \Sigma) \tag{3.9}$$

Following estimation, I predict choice probabilities for each household and choice by simulating R draws, drawing values of δ_{wj} from the distribution $f(\delta|\bar{\delta}, \Sigma)$ and ε_{ij} from the Gumbel distribution.

The marginal effect of a choice variable or profit function variable on participation in an activity can be derived from equation 3.8. The parameters for all options appear in the probability equation for each option. It is not straightforward, ex ante, to predict how occupational choices will vary with profit function variables that appear in profit functions for multiple activities. With the functional forms specified, the marginal effects of each profit and choice variable are allowed to vary across households.

3.4 Data and variables

I estimate the model using household level data from the Tanzanian National Panel Survey, which is part of the LSMS-ISA dataset. These nationally representative, multi-topic and multi-purpose surveys allow for construction of occupational choice, time use, and returns to participation variables. They also include relevant covariates, such as firm and farm inputs and outputs, infrastructure and market access, and household characteristics. I estimate the model using the 2010-11 round of data.

For each household[4], I generate labor supply variables based on individual level, activity-specific time recall variables over the 12 month period preceding the survey date. I then classify households by their corresponding occupational choices $P_i = (P^F, P^E, P^M)$, with participation defined as positive supply of hours by a household member to a given activity and non-participation defined as no supply of labor to the activity. Because I am interested in the annual returns to participation per household, and in the intensive margins of labor supply and occupational choice, I do not differentiate between households who supply different hours of labor to the same activity. If households run an enterprise without any member supplying any labor to it, or if a household operates a farm

[4]In this survey, households are defined as groups of individuals who live together and share meals.

Table 3.1: Tabulation of Occupational Choices

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Rural								
Number HHs	110	871	127	553	119	348	68	251
Share HHs	0.0450	0.356	0.0519	0.226	0.0486	0.142	0.0278	0.103
Per capita consumption, usd	1,054	697.2	1,374	803.5	2,026	669.0	1,269	847.6
(sd)	1,208	484.5	1,340	519.3	1,517	494.3	1,033	605.0
Urban								
Number HHs	96	72	303	123	232	65	189	72
Share HHs	0.0833	0.0625	0.263	0.107	0.201	0.0564	0.164	0.0625
Per capita consumption, usd	2,392	921.4	2,072	1,243	2,554	1,431	2,113	1,201
(sd)	1,821	557.3	1,586	1,134	1,821	1,228	1,629	822.9

without any household member supplying any labor to it, I do not consider this participation from a labor supply perspective. Participation rates and average per capita incomes are tabulated by occupational choice in Table 3.1.

Besides participation, the other dependent variable in the model is returns to participation. The net returns to self employment in a farm enterprise are the gross value of output, including the value of own-consumed or non-marketed farm products, net costs incurred, which include purchase of inputs, non-farm hired labor, machinery, etc. The net returns to self employment in a non-farm enterprise consist of gross firm procedes over the 12-month recall period minus costs incurred. Wage labor net returns consist of the total gross wages earned during the 12-month recall period by household members who worked as laborers during the period. For ease of interpretation, I convert all local currency based measures to constant international dollars using the purchasing power parity conversion factor for private consumption from the World Banks World Development Indicators.

In order to estimate the second stage of the model, it is important to observe all of the first stage covariates not just for households' chosen occupations, but also for non-chosen occupations. The imputed incomes for non-chosen options are reflected in the denominator of each choice probability equation, as depicted in Equation 3.8. Therefore, the first stage estimation of returns to activity participation uses variables that can be observed regardless of the households' occupational choice.

In the agricultural profit functions, the contextual variables are derived from multiple datasets. A general control for agricultural yield potential was created by matching low-technology yield potential estimates from the gridded Global Agro-Ecological Zones dataset with household locations.[5] Yield potential estimates are aggregated across the most country's most widely grown crops[6] using crop area weights generated from crop maps contained in the Harvest Choice dataset.[7] Farm technology is proxied by crop model generated yield predictions for a high technology use scenario and a low technology use scenario. These yield predictions are generated part of the Global Agro-Ecological Zones dataset, using the same crop area weights to create cross-crop measures of technological potential.[8]

Additional agricultural variables from georeferenced sources include mean rainfall during the wettest quarter of the year (from the National Oceanographic and Atmospheric Association), average slope (from the US Geological Survey), soil nutrient retention capacity and workability (from FAO) and the share of land under irrigation (FAO). Using the LSMS-ISA dataset, I generate a locally

[5] http://www.fao.org/nr/gaez/en/

[6] For Tanzania, this list includes maize, paddy rice, cassava, banana, sweet potato, sugar, and cotton.

[7] http://harvestchoice.org/

[8] http://www.fao.org/nr/gaez/en/

smoothed estimate of daily median wages paid per hired farm worker as a proxy for labor prices. The smoothing technique used for median wages and for other variables in this section involves generating a variable at the smallest administrative unit above the household. In the case of Tanzania, this is the ward. If at least three observations of the variable are not available at the ward level, I use the next higher administrative level to generate the statistic. In the case of wages, I use median instead of mean wages in order to reduce the influence of outliers. Land prices are proxied by population density and total area of land owned by the household. Availability of machinery is proxied by the availability of tractors, as described by the locally smoothed rate of tractor use by survey respondents. Table 3.2 contains summaries of the farm profit variables used in profit modeling, tabulated across the eight occupational choices.

The self employment profit variables include the locally smoothed median average annual cost per worker hired by an enterprise, generated from the LSMS-ISA dataset, as a proxy for labor costs. Access to productive capital is proxied by one index of non-agricultural productive assets and another one for agricultural productive assets. Durable household goods like televisions and mattresses are not included in the index. The prevalence of energy inputs is proxied by nighttime light intensity, taken from the Defense Meteorological Satellite data. Table 3.3 summarizes enterprise profit variables by occupational choice.

Wage labor profit variables include locally smoothed median annual returns to wage employment and participation rates for the agricultural, industry, and service sectors. These are meant to proxy for the demand for wage employment. The wage labor variables are summarized by occupational choice in Table 3.4.

Additionally, the maximum educational level attained by any household member is also included, on the premise that the most educated household member is the one most likely to secure wage employment outside of the household.

In all of the profit functions, a common set of demographic and geographic variables is included. This includes a dummy equal to one for urban households, as included in the LSMS dataset. It also includes a dummy for peri-urban households, which are assumed to be those that can travel to a population center of at least 500,000 people within two hours. The household's travel time to the nearest town of 500,000 or above, its network distance to the nearest town of 100,000 or above, and its network distance to the nearest major road are also included. Network distances are generated using maps of transport routes. Household level common demographic variables included in profit functions are the number of household members between the ages of 16 and 65, the age of the household head, a dummy variable the equals one if the household head is female, and the average years of education among household adults, excluding indivduals under the age of 25 who may still be enrolled in school. These common profit variables are summarized by occupational choice in Table 3.5.

Finally, a few additional variables are also included as predictors of households' occupational choices apart from their effects on profits. These include a measure of the household's incoming transfers received from public and private sources. Demographic variables include household size, the share of household members who are dependent (below the age of 15 or over the age of 65), and a dummy that equals one if the household head's father attended school. Finally, I include the average length of the agricultural growing season, which was generated using MODIS global vegetation phenology dataset. These variables are summarized by occupation choice in Table 3.5. Basic summary statistics for all covariates are contained in the appendix in Table **??**.

3.5 Results

I estimate the model using the 2010-11 round of Tanzania survey data. Table 3.6 depicts the marginal effects each variable on annual household returns to participation in farm employment, self employment, and wage employment. The selection variable coefficients are shown in Table 3.7. The first stage model fit is fairly pretty good, with pseudo-R^2 values of 0.32 for farming, 0.20 for self employment, and 0.43 for wage employment.

Table 3.2: Summary statistics for farm variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Yield Potential low (cross-crop ind)	0.475	0.423	0.457	0.435	0.475	0.452	0.465	0.427
(sd)	0.139	0.179	0.121	0.169	0.122	0.169	0.103	0.166
Yield Potential high (cross-crop ind)	0.556	0.506	0.543	0.520	0.551	0.538	0.558	0.519
(sd)	0.159	0.205	0.133	0.187	0.137	0.195	0.125	0.190
Rate improved maize seed use (mean smth)	0.136	0.0822	0.170	0.0973	0.164	0.0878	0.184	0.0998
(sd)	0.203	0.145	0.198	0.162	0.197	0.165	0.193	0.174
Cost hired lbr (med smth, USD/day)	3.826	1.797	4.551	2.074	4.212	1.884	4.844	2.116
(sd)	2.705	1.493	2.740	1.709	2.711	1.534	2.750	1.739
Rate of tractor use (mean smth)	0.0129	0.0452	0.00371	0.0376	0.00875	0.0383	0.00761	0.0301
(sd)	0.0452	0.110	0.0216	0.102	0.0457	0.0885	0.0343	0.0805
Land owned (ha, RIGA)	0.230	1.811	0.166	1.874	0.0427	1.419	0.155	1.723
(sd)	0.620	1.979	0.868	2.375	0.246	1.849	0.674	1.956
Mean precip wettest qrtr (mm, NOAA CPC)	599.8	585.9	578.4	572.1	599.9	579.0	565.7	558.8
(sd)	167.9	193.1	128.8	185.5	146.1	191.9	118.7	165.0
Slope (pct, USGS)	3.798	6.066	3.361	5.081	3.749	5.710	3.046	5.407
(sd)	3.277	6.071	2.638	4.838	3.052	5.286	1.909	5.253
Soil nutrient retention capacity (FAO)	1.330	1.515	1.407	1.544	1.425	1.525	1.440	1.644
(sd)	0.770	0.853	1.109	0.933	0.994	0.719	1.240	1.101
Soil workability (FAO)	1.272	1.682	1.228	1.675	1.208	1.535	1.249	1.731
(sd)	0.902	1.119	1.140	1.184	1.008	0.854	1.269	1.246
Share land irrigated (percent, FAO)	0.00303	0.00765	0.00609	0.00488	0.00385	0.00753	0.00297	0.00436
(sd)	0.0160	0.0319	0.0322	0.0220	0.0249	0.0351	0.0161	0.0212

Table 3.3: Summary statistics for self-employment variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Cost/hired worker (med smth, USD)	1,034	521.3	1,283	549.8	1,354	542.5	1,347	539.2
(sd)	932.8	564.9	1,037	593.2	1,100	581.5	1,081	579.7
Nighttime light ave coverage(DMSP F16)	1,392	34.74	1,778	145.8	1,868	71.59	2,042	207.1
(sd)	1,804	162.4	1,771	629.3	1,741	312.4	1,753	739.2
Financial service available	0.549	0.348	0.579	0.395	0.593	0.429	0.576	0.399
(sd)	0.499	0.477	0.494	0.489	0.492	0.495	0.495	0.491
Productive non-ag asset ind, fact 1	0.675	0.644	1.079	0.885	1.085	0.705	1.279	1.066
(sd)	0.647	0.543	0.813	0.698	0.835	0.589	0.883	0.824
Productive ag-related asset ind, fact 1	0.0234	0.0331	0.0249	0.0397	0.0256	0.0330	0.0290	0.0382
(sd)	0.0155	0.0401	0.0221	0.0598	0.0295	0.0437	0.0399	0.0544
Net returns from ent (USD)	0	0	4,861	2,052	0	0	4,514	1,888
(sd)	0	0	5,525	3,447	0	0	5,370	3,363

Table 3.4: Summary statistics for wage employment variables, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Max educ in hh (yrs)	7.908	6.133	9.416	7.428	10.64	7.421	10.56	8.576
(sd)	5.266	3.906	3.880	3.660	4.444	4.048	3.974	3.720
Nighttime light intensity(DMSP F16)	15.36	1.371	19.81	2.710	20.84	1.848	22.63	3.763
(sd)	17.47	3.081	16.49	6.786	16.03	4.135	15.93	7.888
Returns/ag worker (med smth, USD)	687.7	259.2	726.5	320.0	736.9	343.6	744.6	348.1
(sd)	583.2	333.9	504.2	384.2	568.1	443.5	528.3	421.3
Returns/ind worker (med smth, USD)	1,979	978.0	2,485	1,127	2,462	1,028	2,570	1,173
(sd)	1,302	731.3	1,316	989.4	1,290	848.3	1,412	966.9
Returns/ser worker (med smth, USD)	2,576	1,447	2,874	1,515	3,034	1,382	3,020	1,667
(sd)	2,507	1,785	2,650	1,795	2,623	1,781	2,665	2,188
Partpn in ag emplmt (med smth sh)	0.0562	0.123	0.0263	0.124	0.0308	0.209	0.0321	0.200
(sd)	0.0990	0.134	0.0734	0.139	0.0776	0.181	0.0755	0.192
Partpn in ind emplmt (med smth sh)	0.0740	0.0281	0.0958	0.0435	0.107	0.0612	0.125	0.0621
(sd)	0.0893	0.0661	0.103	0.0789	0.101	0.0991	0.103	0.0956
Partpn in ser emplmt (med smth sh)	0.279	0.116	0.344	0.145	0.441	0.203	0.423	0.243
(sd)	0.193	0.131	0.186	0.146	0.191	0.175	0.175	0.195
Net returns from market (USD)	1,347	113.2	701.0	180.5	4,992	1,701	4,418	2,071
(sd)	3,641	1,008	2,559	1,486	5,170	3,394	5,132	3,753

Table 3.5: Summary statistics for profit choice variables and controls included in profit models for all occupations, by occupational choice

	None	Farm	Self	Farm, Self	Wage	Farm, Wage	Self, Wage	Farm, Self, Wage
Urban	0.466	0.0764	0.705	0.182	0.661	0.157	0.735	0.223
(sd)	0.500	0.266	0.457	0.386	0.474	0.365	0.442	0.417
Peri-urban dummy	0.0388	0.0286	0.0837	0.0607	0.0598	0.0557	0.0428	0.0929
(sd)	0.194	0.167	0.277	0.239	0.238	0.230	0.203	0.291
Hrs travel to nrst town >500k (LSMS-ISA)	5.120	8.288	3.632	7.658	3.793	6.986	2.766	6.790
(sd)	4.770	4.522	4.342	4.624	4.296	4.114	3.521	4.412
Network dist to nrst town >100k (km, LSMS-ISA)	71.94	141.6	49.97	134.2	44.90	131.6	39.00	136.4
(sd)	83.00	94.00	76.51	99.30	72.28	99.87	67.83	101.9
Dist nrst major rd (km, LSMS-ISA)	13.65	24.88	5.836	22.32	6.480	20.59	4.849	16.96
(sd)	23.70	25.37	13.88	25.82	15.73	22.91	12.37	21.20
People per square km, 2005 (ln, HC)	6.092	4.453	6.944	4.660	7.210	4.764	7.296	4.890
(sd)	2.511	1.156	1.996	1.327	1.815	1.153	1.776	1.320
Number hh members 16-65	1.830	2.583	2.581	2.966	2.587	2.976	3.661	3.706
(sd)	1.192	1.593	1.482	1.859	1.574	1.529	2.044	1.947
Age of head	44.96	52.17	41.42	46.75	38.47	47.16	43.37	47.40
(sd)	18.17	16.76	13.60	14.48	12.65	14.90	13.08	14.39
Female head	0.432	0.238	0.323	0.250	0.211	0.228	0.233	0.180
(sd)	0.497	0.426	0.468	0.433	0.408	0.420	0.424	0.384
Yrs educ, adults (ave)	6.918	4.507	8.178	5.894	9.456	5.737	8.871	6.525
(sd)	4.908	3.336	3.424	3.279	4.248	3.644	3.660	3.414
Transfers recvd (USD)	117.2	73.55	67.24	67.25	81.81	83.84	82.88	71.91
(sd)	175.3	137.5	152.5	129.7	184.0	136.2	195.7	143.4
Household size	3.364	5.382	4.421	5.889	4.034	5.872	5.630	6.755
(sd)	2.261	2.946	2.509	3.390	2.634	2.830	2.954	3.213
HH dependent share	0.303	0.414	0.247	0.359	0.176	0.347	0.215	0.323
(sd)	0.334	0.268	0.242	0.233	0.209	0.227	0.199	0.206
Head's father attended school	0.461	0.310	0.665	0.395	0.701	0.429	0.642	0.498
(sd)	0.500	0.463	0.472	0.489	0.459	0.495	0.480	0.501
Years educ, head	7.083	4.536	8.281	5.914	9.744	5.690	8.844	6.570
(sd)	5.252	3.918	3.924	3.900	4.615	4.187	4.540	4.393
Ave length of season (days, MOD12Q2)	182.4	174.0	183.8	172.7	184.1	179.2	186.4	177.6
(sd)	15.85	24.98	16.68	25.02	17.15	26.03	15.69	25.62

Table 3.6: Marginal effects of profit function variables on returns to participation in farming, self employment, and wage employment (in USD per year). These are based on the Generalized Leontief profit function specification estimated using a Heckman selection model. The selection parameter estimates are shown in Table 3.7.

	Farm Margins at means	(SE)	Self Emplm't Margins at means	(SE)	Wage Emplm't Margins at means	(SE)
Urban (sq rt)	135.842	105.371	-710.422	952.864	-323.012	746.940
Peri-urban dummy (sq rt)	-11.949	678.455	-5,809.333	4,826.420	1,515.902	3,555.486
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-87.897	53.434	-987.280	531.530	92.147	328.373
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	14.805	8.933	-28.612	91.065	-64.368	59.958
Dist nrst major rd (km, LSMS-ISA) (sq rt)	15.000	18.173	77.699	234.559	-237.193*	114.297
People per square km, 2005 (ln, HC) (sq rt)	-204.729	122.090	1,013.498	1,526.430	348.260	897.252
Number hh members 16-65 (sq rt)	360.126**	64.025	804.815	750.843	508.537	395.601
Age of head (sq rt)	43.922	25.068	-419.253	273.632	42.620	145.741
Female head (sq rt)	-54.241	56.824	-201.497	538.188	-797.558*	341.084
Yrs educ, adults (ave) (sq rt)	66.084	35.784	-42.117	379.546	-154.296	624.761
Yield Potential low (cross-crop ind) (sq rt)	-185.002	460.278				
Yield Potential high (cross-crop ind) (sq rt)	-94.349	436.320				
Rate improved maize seed use (mean smth) (sq rt)	-184.248	136.459				
Cost hired lbr (med smth, USD/day) (sq rt)	-121.834	67.761				
Rate of tractor use (mean smth) (sq rt)	407.344	323.785				
Land owned (ha, RIGA) (sq rt)	383.472**	40.049				
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	30.411*	12.090				
Slope (pct, USGS) (sq rt)	59.790	43.512				
Soil nutrient retention capacity (FAO) (sq rt)	-217.076	119.507				
Soil workability (FAO) (sq rt)	216.567	130.760				
Share land irrigated (percent, FAO) (sq rt)	2,556.904**	899.619				
Cost/hired worker (med smth, USD) (sq rt)			24.064	28.378		
Nighttime light ave coverage(DMSP F16) (sq rt)			-30.658	44.336		
Financial service available (sq rt)			1,003.185	513.653		
Productive non-ag asset ind, fact 1 (sq rt)			3,171.157**	1,041.906		
Productive ag-related asset ind, fact 1 (sq rt)			-4,581.303	7,078.165		
Max educ in hh (yrs) (sq rt)					1,980.455**	586.246
Nighttime light intensity(DMSP F16) (sq rt)					49.682	163.018
Returns/ag worker (med smth, USD) (sq rt)					-16.314	35.770
Returns/ind worker (med smth, USD) (sq rt)					-2.169	17.224
Returns/ser worker (med smth, USD) (sq rt)					12.166	11.823
Partpn in ag emplmt (med smth sh) (sq rt)					-313.959	1,021.808
Partpn in ind emplmt (med smth sh) (sq rt)					-359.235	1,006.978
Partpn in ser emplmt (med smth sh) (sq rt)					254.652	803.682
N	3,599		3,599		3,599	
N (non-censored)	2407		447		1342	
Mean profits (USD/yr)	1123.93		3341.16		3767.53	
R2 (adj)	0.327		0.256		0.362	

* $p < 0.05$; ** $p < 0.01$

Table 3.7: Selection variable coefficients for returns to participation by activity (Generalized Leontief profit function specification with selection). Parameters were estimated using a Heckman Selection Model. The profit accompanying function parameters are shown in Table 3.6.

	Farm Select	(SE)	Self Emplm't Select	(SE)	Wage Emplm't Select	(SE)
Transfers recvd (USD)	-0.001**	0.000	-0.000*	0.000	0.000	0.000
Household size	0.099**	0.010	0.085**	0.009	0.053**	0.008
HH dependent share	0.623**	0.106	-0.124	0.126	-0.797**	0.101
Head's father attended school	-0.035	0.055	0.153*	0.062	0.112*	0.048
Years educ, head	-0.063**	0.006	0.009	0.007	0.049**	0.006
Urban	-1.391**	0.055	0.186**	0.063	0.558**	0.051
Peri-urban dummy	-0.619**	0.103	0.323**	0.114	0.367**	0.098
Ave length of season (days, MOD12Q2)	-0.011**	0.001	-0.002*	0.001	0.004**	0.001
Lambda	-390.54		-977.66		-1,697.31	
Sigma	800.66		3,568.35		3,804.23	
P value comparison test	0.00		0.00		0.00	
N	3,599		3,599		3,599	
N (censored)	1,192		3,152		2,257	
N (non-censored)	2407		447		1342	

* $p < 0.05$; ** $p < 0.01$

For farming, the variables that have a marginal effect on annual profits that is significantly different from zero, after controlling for selection, are: household size (positive), land owned (positive), precipitation during the wettest quarter of the year (positive), hired labor costs (negative), and soil nutrient retention (negative). For self employment, the marginal effects that are significantly different from zero include: non-agricultural productive assets index (positive), peri-urban dummy, with rural as the base case (negative), and the age of the household head (negative). For wage employment , the marginal effects that are significant include the years of education by the most educated household member (positive), the average local wages for a service sector worker (positive), the distance to the nearest major road (negative), and a dummy for female headed households, with male-headed households as the base case (negative). I have included tables in the appendix that depict all of the coefficients for each variable interaction, in matrix form (Table **??** for farming, Table **??** for self employment, and Table **??** for wage employment).

The parameter estimates for the second stage occupational choice model are depicted in Table 3.8. The income parameter (α), and the parameters that describe the multi-variate normal distribution of the alternative-specific coefficients are shown. The parameter estimates are consistent with scale heterogeneity across choices.

In Table 3.9, I show the average marginal effect of each profit function variable on the probability of selecting each choice, along with the standard deviation of the marginal effects. The profit function variables affect occupational choice through income effects. They are calculated by differentiating the closed form solution of the probability of participating in an occupation with respect to the each x variable. This expression is evaluated at each data point, drawing simulated δ coefficients using the estimated parameters of the multivariate normal distribution. Similarly in Table 3.10, I show the average marginal effect of each selection variable on the probability of choosing each occupational choice.

The choice shares predicted by the model match very closely with the participation shares observed in the data. Figure **??** shows the actual particpation shares compared with the predicted shares, along with a box plot of the 5th to the 95th percentiles of prediction probabilities. There is good fit across the entire estimation sample and within specific subsamples, of the dataset. Figure **??** shows a scatterplot of predicted probabilities onto actual participation shares for sixteen different subsets of the population. There is a very good fit between predicted probabilities and actual choice shares within all groups for which fit was checked. Next, I performed a validation exercise by estimating the model on a subset of data, randomly dropping one fifth of the sample enumeration areas. Figure **??** depicts a comparison between predicted probabilities and actual choice shares for the enumeration areas not used in model estimation, showing a fairly close fit.

Occupational choices do not vary greatly over agroclimatic potential, as characterized by a cross-crop index of medium-technology yield levels (Figure 3.2). Self and wage employment are much more common in high population density areas than in low population density areas. And farming is much more common in low population density areas than in high population density areas (Figure 3.3). This is consistent with high population density areas featuring larger markets for those operating household non-farm enterprises. High population density areas are also more likely to have surplus labor supply and more wage labor employment opportunities. Households located nearer to population centers of at least 100k people are more likely to participate in wage and self employment than are households located farther from these population centers (Figure 3.4). Those located further away from population centers are more likely to farm, or to farm in addition to participating in wage or self employment.

Table 3.8: Second stage occupational choice model parameter coefficients. The base case is non-participation in all activities. The left column shoes coefficients in the utility function for income and income squared. The matrix below it shows alternate-specific coefficients for each of the selection variables, the mean random coefficient for each alternative, and the standard deviation for each alternative (the diagonal of the variance-covariance matrix desctibing the multi-variate normal distribution of δ.

	MSL						
Income per capita sq rt. (predicted '000)	0.414*						
	(0.175)						
Income per capita sq rt. X urban	-0.290						
	(.065)						
	farm	self	farm,self	wage	wage,farm	wage,self	farm,wage,self
Transfers recvd (USD)	-2.199**	-1.231*	-2.504**	-0.409	-1.622**	-0.775	-2.021**
Household size	0.477**	0.335**	0.566**	0.293**	0.568**	0.551**	0.673**
HH dependent share	0.404	-0.904*	-0.339	-2.072**	-0.591	-2.550**	-1.371**
Head's father attended school	-0.016	0.549**	0.126	0.540**	0.295	0.466*	0.481*
Years educ, head	-0.091**	0.002	-0.034	0.074**	-0.052*	0.015	-0.021
Urban	-1.194**	1.266**	-0.239	0.791**	-0.403	1.373**	0.081
Peri-urban	-1.370**	1.736**	-0.377	1.252**	-0.790	0.985	-0.013
Ave length of season	-0.019**	-0.005	-0.024**	-0.004	-0.012**	0.001	-0.017**
δ (mean)	3.054**	-0.614	2.880**	-0.992	0.375	-3.182**	0.188
δ (sd)	(0.791)	(0.884)	(0.881)	(0.920)	(0.846)	(1.045)	(0.896)
N	3599						
Log-likelihood	-5773.68						
Pseudo R2	.1225						
AIC	907.37						
BIC	1489.08						

Table 3.9: Average marginal effect of each profit function variable on the probability of participating in each occupation. The standard deviation of each average marginal effect estimate is shown below in parentheses.

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
Common Profit Function Variables									
Urban	-0.00275	-0.177	0.0798	0.0164	0.0199	-0.00865	0.0542	0.0252	0.00141
	(0.0446)	(0.0933)	(0.0635)	(0.0742)	(0.0405)	(0.0385)	(0.0493)	(0.0436)	(0.0952)
Peri-urban	0.000879	-0.168	0.112	-0.0183	0.0636	-0.0217	0.0186	0.0245	0.00160
	(0.0602)	(0.0875)	(0.0944)	(0.0818)	(0.0648)	(0.0512)	(0.0384)	(0.0504)	(0.104)
Travel time (>500k)	0.000359	0.0000617	-0.00260	-0.000973	0.00210	0.000813	-0.000198	0.0000514	-0.0000763
	(0.00189)	(0.00510)	(0.0111)	(0.00608)	(0.00885)	(0.00450)	(0.00115)	(0.000869)	(0.00632)
Network Dist (>100k)	-0.00000177	0.0000306	-0.0000108	-0.0000136	0.0000103	-0.00000314	-0.000000205	-0.00000845	0.000000410
	(0.0000396)	(0.000114)	(0.000118)	(0.0000911)	(0.000128)	(0.0000793)	(0.0000197)	(0.0000247)	(0.0000899)
Dist to major rd	0.000118	0.000464	0.000185	0.000411	-0.000457	-0.000412	-0.000260	-0.0000548	-0.00000801
	(0.00196)	(0.0102)	(0.00240)	(0.00494)	(0.00306)	(0.00593)	(0.00124)	(0.00354)	(0.00508)
Pop density	0.0000761	-0.0000872	-0.000549	0.00310	-0.000483	-0.00107	-0.000144	-0.000620	0.0000284
	(0.0186)	(0.0374)	(0.00696)	(0.219)	(0.0507)	(0.0718)	(0.0101)	(0.0353)	(0.0884)
HH indep pop	-0.0164	0.0521	-0.00224	-0.0170	-0.00151	-0.0167	-0.000778	-0.00409	-0.0000129
	(0.165)	(0.414)	(0.0950)	(0.185)	(0.0215)	(0.153)	(0.00857)	(0.0527)	(0.185)
Age hh head	-0.0000162	0.0000402	-0.0000365	-0.0000901	0.0000411	0.0000635	-0.00000387	-0.00000132	0.000000378
	(0.000222)	(0.000407)	(0.000249)	(0.000491)	(0.000214)	(0.000215)	(0.0000544)	(0.000121)	(0.000287)
Female head dummy	0.00120	0.00120	-0.000352	0.000657	0.000108	-0.00161	-0.000548	-0.000822	-0.0000964
	(0.00490)	(0.00866)	(0.00664)	(0.00821)	(0.00457)	(0.00579)	(0.00164)	(0.00282)	(0.00598)
Ave hh educ (yrs)	0.0398	0.213	0.0271	0.107	-0.0382	-0.253	-0.0178	-0.0738	-0.00230
	(0.173)	(0.686)	(0.146)	(0.340)	(0.186)	(0.861)	(0.0780)	(0.224)	(0.458)
Farm Profit Function Variables									
Low tech yield potential	-0.000520	-0.00592	-0.0000453	-0.000662	0.000365	0.00436	0.000191	0.00183	-0.0000105
	(0.0154)	(0.188)	(0.00753)	(0.0478)	(0.0140)	(0.101)	(0.00423)	(0.0477)	(0.0815)
High tech yield potential	0.00175	0.000149	0.000462	0.00112	0.000170	-0.00246	0.0000616	-0.000710	-0.0000336
	(0.0138)	(0.129)	(0.00597)	(0.0343)	(0.00959)	(0.0700)	(0.00328)	(0.0328)	(0.0563)
Maize tech use (share)	0.0146	-0.0366	0.0345	-0.0245	0.0123	-0.00404	0.0123	0.00635	0.00115
	(0.729)	(1.852)	(0.380)	(0.651)	(0.277)	(0.480)	(0.143)	(0.247)	(0.784)
Farm wages (cost/day)	0.00000548	-0.000144	0.00000969	0.0000299	0.0000387	0.0000178	0.0000196	0.0000495	0.00000349
	(0.000897)	(0.00224)	(0.000481)	(0.000837)	(0.000342)	(0.000771)	(0.000147)	(0.000300)	(0.000976)
Tractor use (share)	-0.354	1.424	-0.0931	-0.236	-0.203	-0.181	-0.137	-0.283	0.00981
	(1.948)	(6.339)	(1.797)	(2.450)	(1.157)	(1.652)	(0.511)	(0.860)	(2.781)
Land owned (ha)	-0.0311	0.0944	-0.0225	-0.00146	-0.0216	-0.00108	-0.0107	-0.00685	0.00158
	(0.177)	(0.409)	(0.0745)	(0.160)	(0.0746)	(0.0950)	(0.0280)	(0.0437)	(0.179)
Rainfall	-0.000000755	0.00000383	-0.00000102	-0.000000347	-0.000000770	-1.90e-08	-0.000000470	-0.000000294	5.84e-08
	(0.00000604)	(0.0000207)	(0.00000342)	(0.00000906)	(0.00000299)	(0.00000526)	(0.00000139)	(0.00000263)	(0.00000890)

(*Continued on next page*)

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
(*Continued from previous page*)									
Slope	-0.0000353	0.000192	-0.0000293	-0.0000232	-0.0000373	-0.0000168	-0.0000222	-0.0000245	0.00000224
	(0.000398)	(0.00257)	(0.000238)	(0.000566)	(0.000224)	(0.000775)	(0.000127)	(0.000561)	(0.00103)
Soil nutrients	0.0307	-0.157	0.0378	0.0219	0.0229	0.00803	0.0127	0.0128	-0.00291
	(0.253)	(1.128)	(0.316)	(0.434)	(0.164)	(0.265)	(0.0679)	(0.127)	(0.478)
Soil workability	-0.0280	0.157	-0.0397	-0.0224	-0.0235	-0.00619	-0.0127	-0.0127	0.00299
	(0.221)	(1.172)	(0.360)	(0.446)	(0.179)	(0.280)	(0.0701)	(0.128)	(0.498)
Irrigation (share)	-2.100	11.31	-4.033	-0.307	-2.506	0.409	-1.378	-0.844	0.180
	(10.30)	(35.56)	(16.94)	(14.17)	(8.618)	(17.63)	(3.581)	(4.815)	(17.72)
Self Employment Profit Function Variables									
Cost per firm worker	-0.000000968	-0.00000453	0.00000163	0.00000521	-0.00000116	-0.00000180	0.000000389	0.00000102	4.41e-08
	(0.00000458)	(0.0000170)	(0.00000748)	(0.0000209)	(0.00000428)	(0.00000665)	(0.00000234)	(0.00000424)	(0.0000111)
Nighttime lights	-0.0000372	-0.000318	0.0000550	0.000338	-0.0000253	-0.000131	0.0000199	0.0000920	0.00000231
	(0.000188)	(0.000784)	(0.000247)	(0.000893)	(0.000142)	(0.000324)	(0.0000779)	(0.000225)	(0.000504)
Financial services available	0.000142	0.0000488	-0.000200	-0.0000363	0.0000105	-0.000104	0.00000308	0.0000677	-0.0000174
	(0.00189)	(0.00607)	(0.00368)	(0.00679)	(0.00201)	(0.00231)	(0.000713)	(0.00157)	(0.00383)
Asset index (non-ag)	0.00280	-0.00501	0.000742	0.0182	-0.00851	-0.00753	0.00176	0.000132	0.000207
	(0.569)	(0.668)	(0.407)	(1.188)	(0.121)	(0.190)	(0.0466)	(0.175)	(0.551)
Asset index (ag)	-0.0236	-0.0906	0.0421	0.0975	-0.0201	-0.0295	0.00696	0.0170	0.00165
	(0.0954)	(0.425)	(0.143)	(0.454)	(0.0605)	(0.105)	(0.0224)	(0.0595)	(0.244)
Wage Employment Profit Function Variables									
Max educ in hh (yrs)	-0.0226	-0.119	-0.0111	-0.0483	0.0192	0.130	0.00958	0.0397	0.00123
	(0.0993)	(0.387)	(0.0599)	(0.159)	(0.0926)	(0.455)	(0.0417)	(0.122)	(0.244)
Nighttime lights	-0.00323	-0.0238	-0.000162	-0.0128	0.00445	0.0245	0.00127	0.00900	0.000132
	(0.0211)	(0.0619)	(0.0246)	(0.0396)	(0.0402)	(0.0665)	(0.00944)	(0.0295)	(0.0440)
Wages (ag worker)	-0.000000355	0.000000220	-0.000000282	0.000000238	0.000000566	-0.000000283	0.000000129	-0.000000171	2.82e-08
	(0.00000444)	(0.0000145)	(0.00000558)	(0.00000910)	(0.00000861)	(0.0000165)	(0.00000186)	(0.00000649)	(0.00000979)
Wages (ind worker)	0.000000654	0.00000231	0.000000454	0.00000153	-0.00000109	-0.00000238	-0.000000271	-0.00000105	-2.24e-08
	(0.00000330)	(0.00000872)	(0.00000314)	(0.00000588)	(0.00000575)	(0.00000928)	(0.00000127)	(0.00000415)	(0.00000609)
Wages (service worker)	-0.000000238	-0.000000674	-0.000000382	-0.000000586	0.000000597	0.000000745	0.000000198	0.000000349	1.62e-08
	(0.00000178)	(0.00000360)	(0.00000168)	(0.00000238)	(0.00000281)	(0.00000408)	(0.000000690)	(0.00000171)	(0.00000265)
Participation share (ag)	-0.242	-0.413	-0.532	-0.269	0.727	0.466	0.206	0.157	0.0276
	(0.954)	(1.572)	(2.031)	(1.147)	(2.611)	(2.022)	(0.531)	(0.764)	(1.686)
Participation share (ind)	0.435	2.122	0.294	1.255	-0.724	-2.290	-0.185	-0.861	-0.0233
	(1.678)	(4.157)	(1.410)	(2.369)	(2.726)	(4.889)	(0.589)	(1.583)	(3.114)
Participation share (ser)	-0.0289	-0.0584	-0.000996	0.0263	0.00570	0.0801	-0.00265	-0.0228	0.00160
	(0.432)	(1.833)	(0.361)	(0.914)	(0.618)	(2.216)	(0.179)	(0.644)	(1.159)

Table 3.10: Average marginal effect of each selection variable on the probability of participating in each occupation variable margins. The standard deviation of each average marginal effect estimate is shown below in parentheses.

	None	Farm	Self	Farm_Self	Wage	Farm_Wage	Self_Wage	Farm_Self_Wage	Total
Income transfers ('000 USD)	0.0706	-0.0565	-0.00942	-0.110	0.0649	0.0298	0.0264	-0.0159	-1.35e-10
	(0.0681)	(0.0482)	(0.0643)	(0.0510)	(0.0552)	(0.0415)	(0.0341)	(0.0280)	(0.0765)
HH Size	-0.0217	-0.00683	-0.00717	0.0119	-0.00903	0.00751	0.0105	0.0148	7.23e-11
	(0.0206)	(0.0125)	(0.00998)	(0.00745)	(0.00935)	(0.00506)	(0.0161)	(0.00856)	(0.0172)
HH Dependency share	0.0470	0.174	0.0304	0.0137	-0.0820	-0.0178	-0.0940	-0.0714	1.00e-10
	(0.0694)	(0.0946)	(0.0834)	(0.0613)	(0.0775)	(0.0421)	(0.0944)	(0.0597)	(0.111)
Head's father school (dum)	-0.0166	-0.0442	0.0200	-0.0121	0.0146	0.0110	0.00508	0.0221	-2.65e-12
	(0.0221)	(0.0242)	(0.0167)	(0.0155)	(0.0132)	(0.0127)	(0.00574)	(0.0171)	(0.0271)
Head's education (years)	0.000580	-0.0101	-0.000103	0.00192	0.00647	-0.000929	0.000697	0.00151	1.35e-12
	(0.00228)	(0.00551)	(0.00344)	(0.00413)	(0.00592)	(0.00198)	(0.00178)	(0.00228)	(0.00575)
Urban	-0.00275	-0.177	0.0798	0.0164	0.0199	-0.00865	0.0542	0.0252	0.00141
	(0.0446)	(0.0933)	(0.0635)	(0.0742)	(0.0405)	(0.0385)	(0.0493)	(0.0436)	(0.0952)
Peri-Urban	0.000879	-0.168	0.112	-0.0183	0.0636	-0.0217	0.0186	0.0245	0.00160
	(0.0602)	(0.0875)	(0.0944)	(0.0818)	(0.0648)	(0.0512)	(0.0384)	(0.0504)	(0.104)
Growing season length	0.000513	-0.000536	0.000260	-0.00141	0.000286	0.000439	0.000608	-0.000164	-4.19e-13
	(0.000528)	(0.000485)	(0.000507)	(0.000631)	(0.000369)	(0.000505)	(0.000606)	(0.000327)	(0.000814)

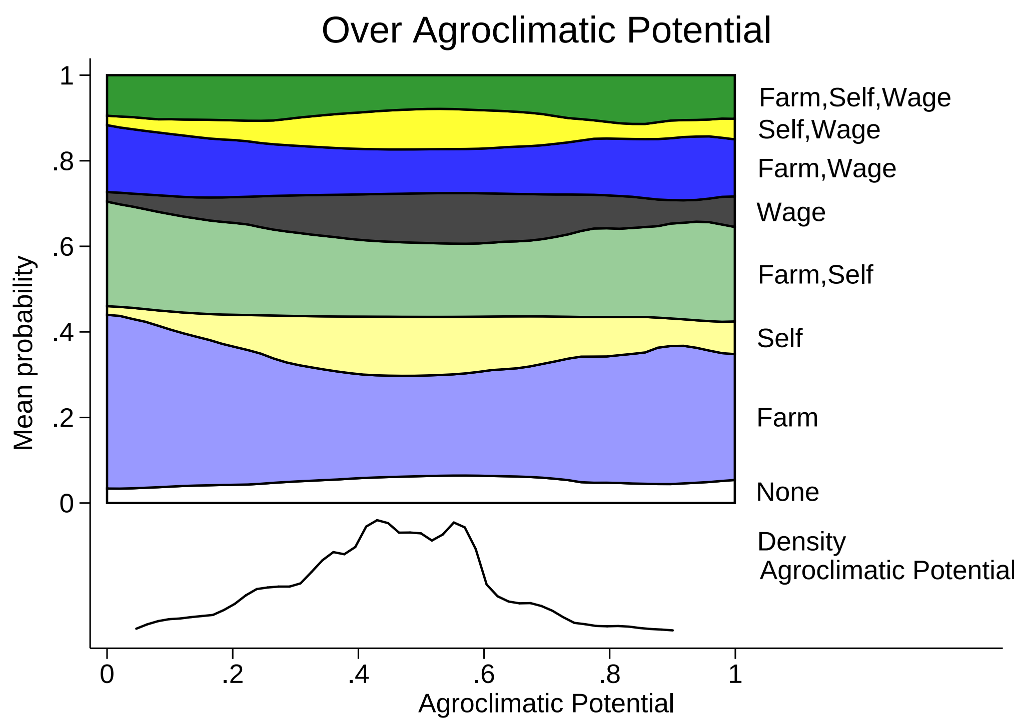
3.6 Policy Simulations

Understanding the sensitivity of occupational choice to labor productivity growth in different sectors has important implications for prioritizing between and targeting delivery of development interventions. A major pathway by which technology-led agricultural labor productivity improvement has resulted in poverty reduction, historically, has been through the eventual reallocation of labor out of agriculture. These pathways are very context-specific, however, and depend on returns to different income-earning opportunities that households face in lieu of, or in addition to, farming.

I simulate three stylized labor productivity shocks in order to understand how these interventions are likely to affect welfare, and the relative importance of shifting occupational choices vis a vis within-sector welfare gains. The first relates to farm labor productivity. I double each households' imputed measure of farm labor productivity. In the second simulation, I double each households' imputed measure of self employment productivity. And, in the third simulation, I double each households' imputed measure of wage labor productivity. For each of the simulations, I also run variants targeting households with low and high agroclimatic potential, low and high population density, and low and high levels of market access, as measured by the travel distance to the nearest population center of at least 500,000 people. With respect to each context variable, "low" and "high" are defined as below and above the median value observed in the sample, respectively.

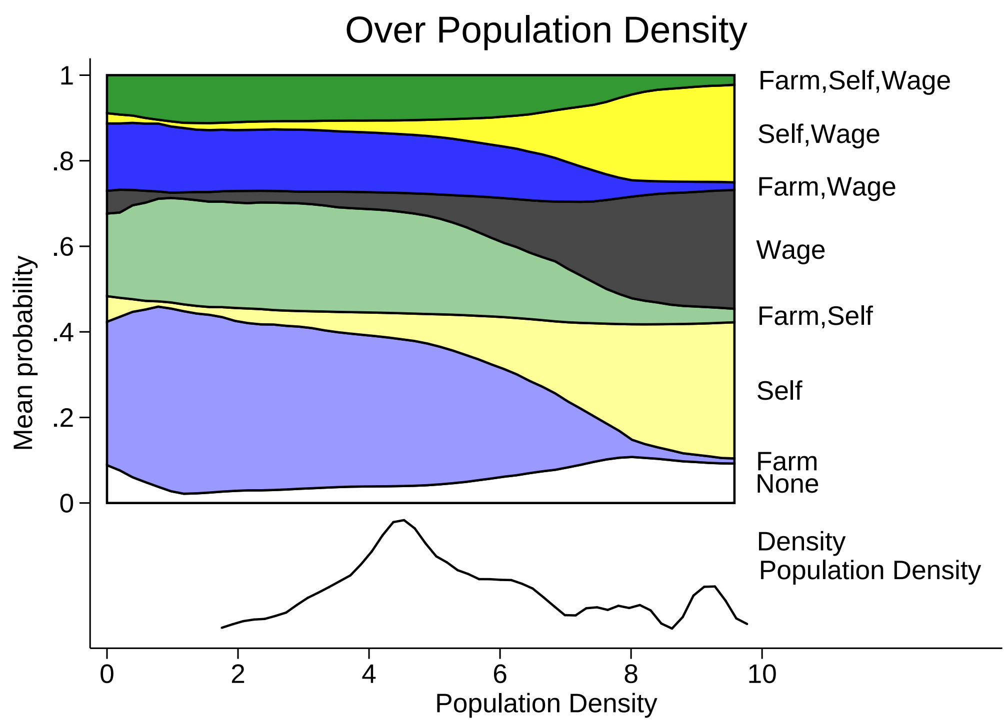
For each simulation, I generate a new values of returns to participating in each occupational choice. These are based on the first stage prediction of labor

Figure 3.2: Conditional probability of participation in each occupation over agroclimatic potential. The probability of each choice corresponding with a specific level of agroclimatic potential is the vertical distance between the line above and the line below the area labeled with that choice.



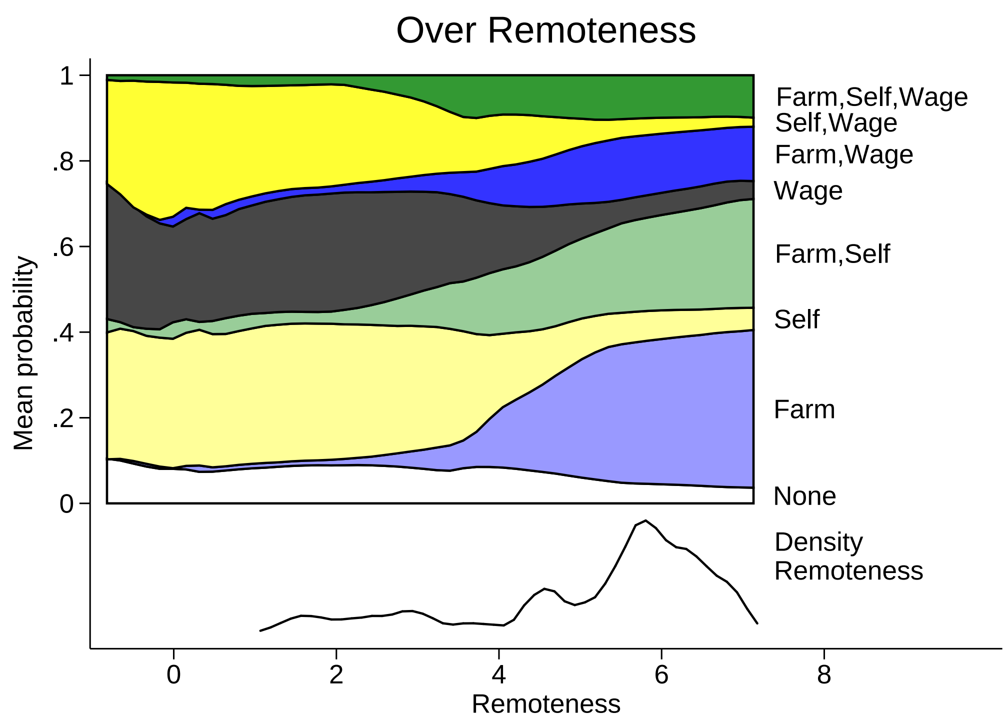
productivity for each household, which are conditioned on all of the variables that are arguments in the profit function for each activity. Using the newly simulated imputed income for each occupational choice, I predict new choice probabilites for each household. I compare the probability of each occupational choice between the baseline and the simulated policy intervention. Then, I compare the baseline welfare with simulated welfare, decomposing welfare changes into those that take place along the intensive margin of participation (without changing occupational choice) and those that take place along the extensive margin of participation (due to a change in occupational choice).

Figure 3.3: Conditional probability of participation in each occupation over population density. The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.



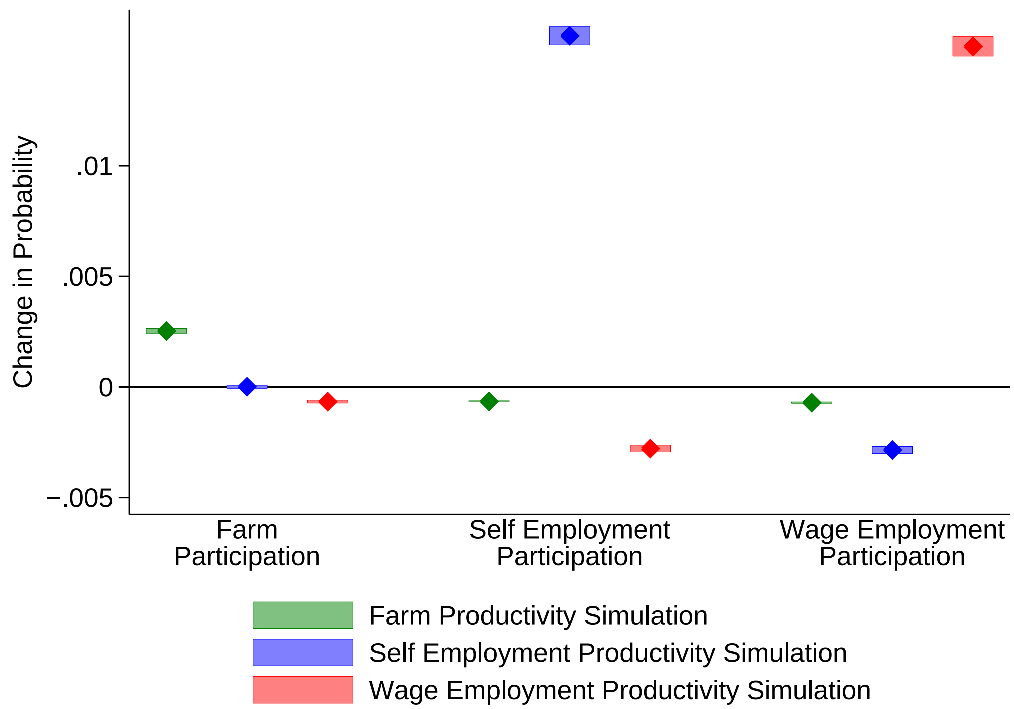
For each of the three simulations, Figure 3.5 shows a box plot of the probability difference in participation in farming, self employment, and wage employment, respectively. Farming participation is not very responsive to any of the simulations. Houses, on average, face a small increase in the probability of participating in farming when farm labor productivity is doubled. Increased self and wage labor productivity are associated with very small decreases in farming participation. Self employment participation increases by 1.5 percentage points when self employment income doubles, and it decreases by less than half of a percentage point when wage labor productivity doubles. Similarly, wage labor participation increases by 1.5 percentage points when wage labor productivity

Figure 3.4: Conditional probability of participation in each occupation over remoteness (log of distance in km to the nearest population center of >500k). The probability of each choice corresponding with a specific population density is the vertical distance between the line above and the line below the area labeled with that choice.



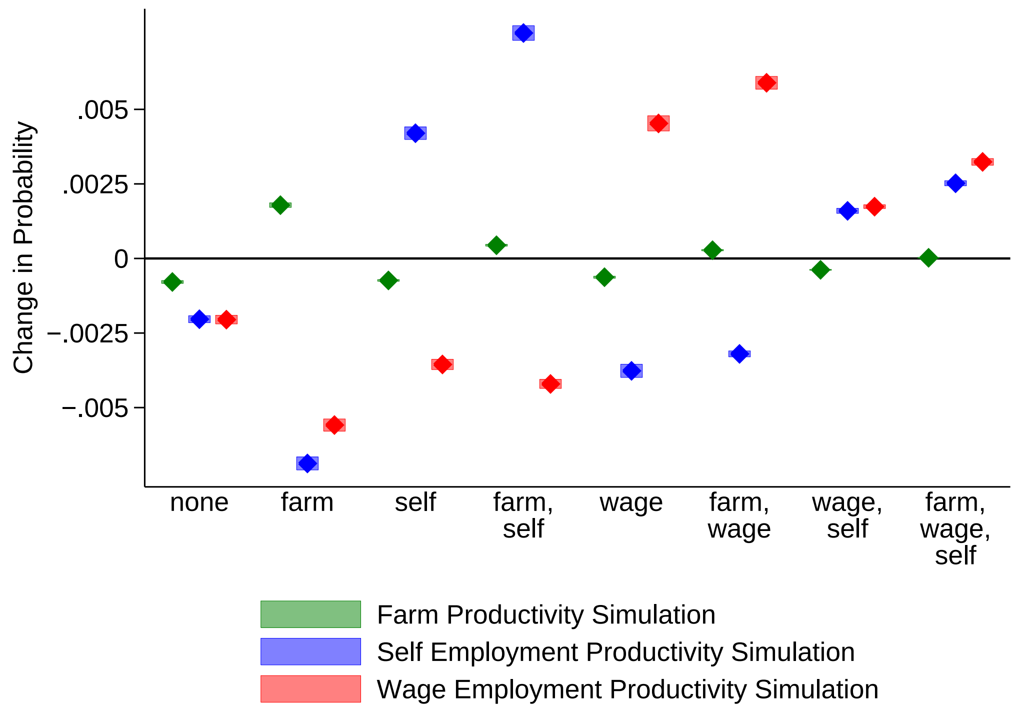
doubles, and self employment decreases by less than half of a percentage point. Households tend to respond to productivity shocks by entering into the activity whose productivity was shocked without exiting from baseline activities in which the household participated (Figure 3.6). The self employment simulation is associated with households that only farm adding self employment and households that do nothing adding self employment. Similarly, the wage employment simulation is associated with households that only farm adding wage employment, and households that participate in farming and self employment adding wage employment (Figure 3.7).

Figure 3.5: Difference in probability of participation in farming, self employment and wage employment across labor productivity simulations. In the farm simulation (depicted in green), farm labor productivity was doubled. In the self employment simulation (depicted in blue), self employment labor productivity was doubled. In the wage employment simulation (depicted in red), wage labor productivity was doubled. The diamonds show the mean difference in probability that households participate in each activity for each simulation, and the bars above and blow each diamond depict the 95% confidence intervals around the population mean.



The expected welfare effects associated with each simulation closely mirror the probabilities of falling into each category, as depicted in Figure 3.8. The self and wage labor producivity simulations are associated with higher average welfare gains in the population than the farm productivity simulation even though farming participation rates are much higher than self and wage employment participation rates, mostly because self and wage employment comprise a high

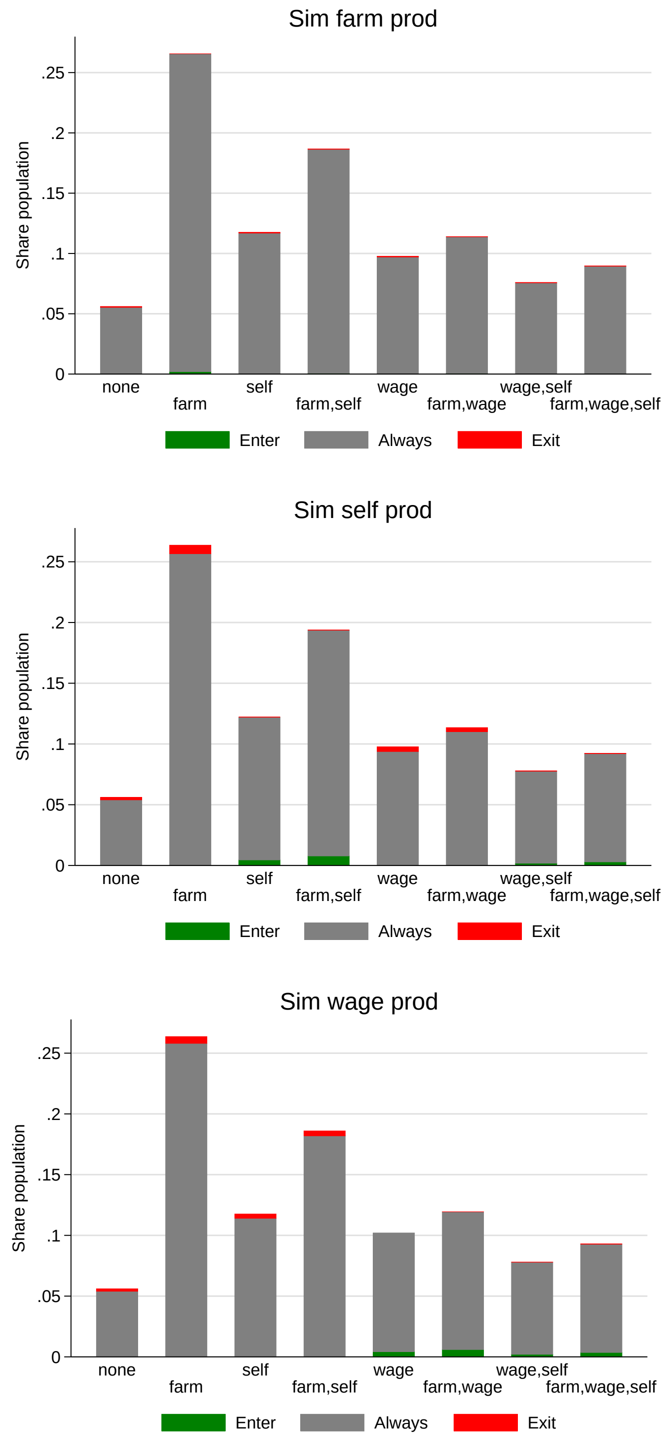
Figure 3.6: Difference in probability of participating in each occupational choice (combination of farming, self employment, and wage employment) across labor productivity simulations described in Figure 3.5.



share of incomes for households that do participate in them.

The probability of changing occupation is examined more closely in different contexts in Figures 3.9, 3.11, and 3.13. The relative welfare gains are examined more closely in different contexts in Figures 3.10, 3.12, and 3.14. In the places with better market access, households facing self employment productivity shocks are more likely to exit wage labor, and households facing wage labor productivity shocks are more likely to exit self employment. This result could arise from lower underemployment or greater returns to specialization in good market access areas. A closer look at the welfare gains to productivity shocks shows that the farm productivity simulation has the greatest impact on

Figure 3.7: The probability that each household falls into the category of entering, always participating in, or exiting each each occupational choice (combination of farming, self employment, and wage employment) across the three labor productivity simulations described in Figure 3.5.



farming households in remote areas (Figure 3.10). The impacts of self and wage employment productivity shocks, on the other hand, are greatest in places with good market access.

As with good market access areas, households with higher population density tended to be more likely to exit wage labor as they entered self employment due to a self employment productivity shock, or to exit self employment as they enter wage labor due to a wage labor productivity shock (Figure 3.11). The expected welfare gains to farming productivity improvement are higher in low population density areas, while the gains to wage and self employment productivity shocks are higher in high population density areas (Figure 3.12). Welfare gains and the probability of changing occupations do not vary over agroclimatic potential (Figures 3.13 and 3.14).

Together, these findings suggest that improved agricultural productivity has a very important role to play, especially for households in remote areas with low population density. They also suggest that increased labor productivity in wage labor and self employment is likely to pull farming households into those activities. However, entry into wage and self employment is not associated with labor exits from agriculture. Income diversification at the household level is likely to play a critical first step in the structural transformation process, and in the eventual shift of labor out of agriculture.

Figure 3.8: Welfare effects (% change in utility between simulated policy intervention and the baseline) of each policy simulation. For each of the three activities – farming, self employment and wage employment – the average welfare gains are decomposed into welfare gains from entry into the activity (when non-participating households shift to participation), exit from the activity (when participating households cease participation), and gains that occur within the participation margin (households participate before and after the intervention).

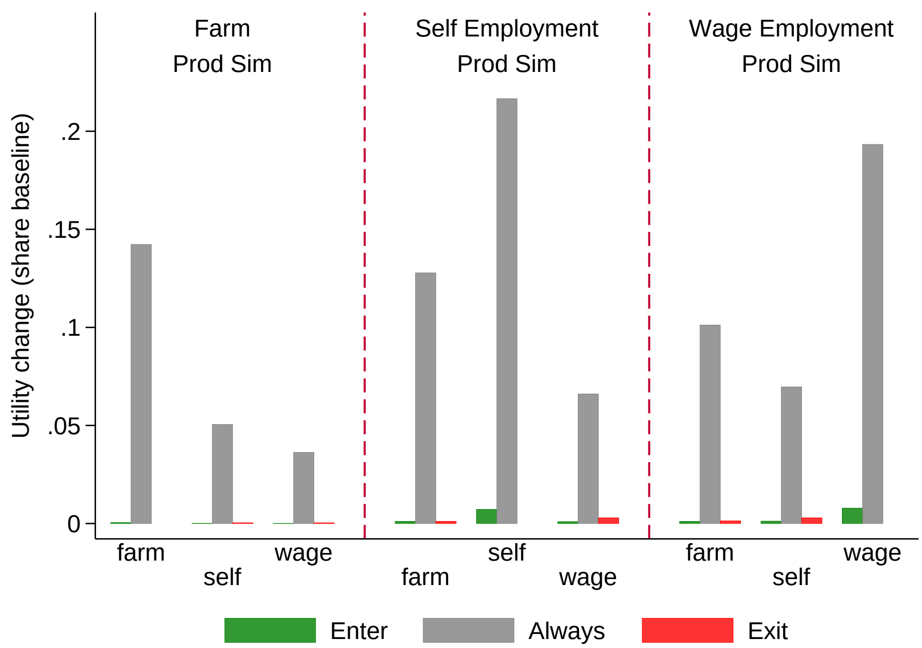


Figure 3.9: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on remoteness. The density of remoteness is shown underneath each regression.

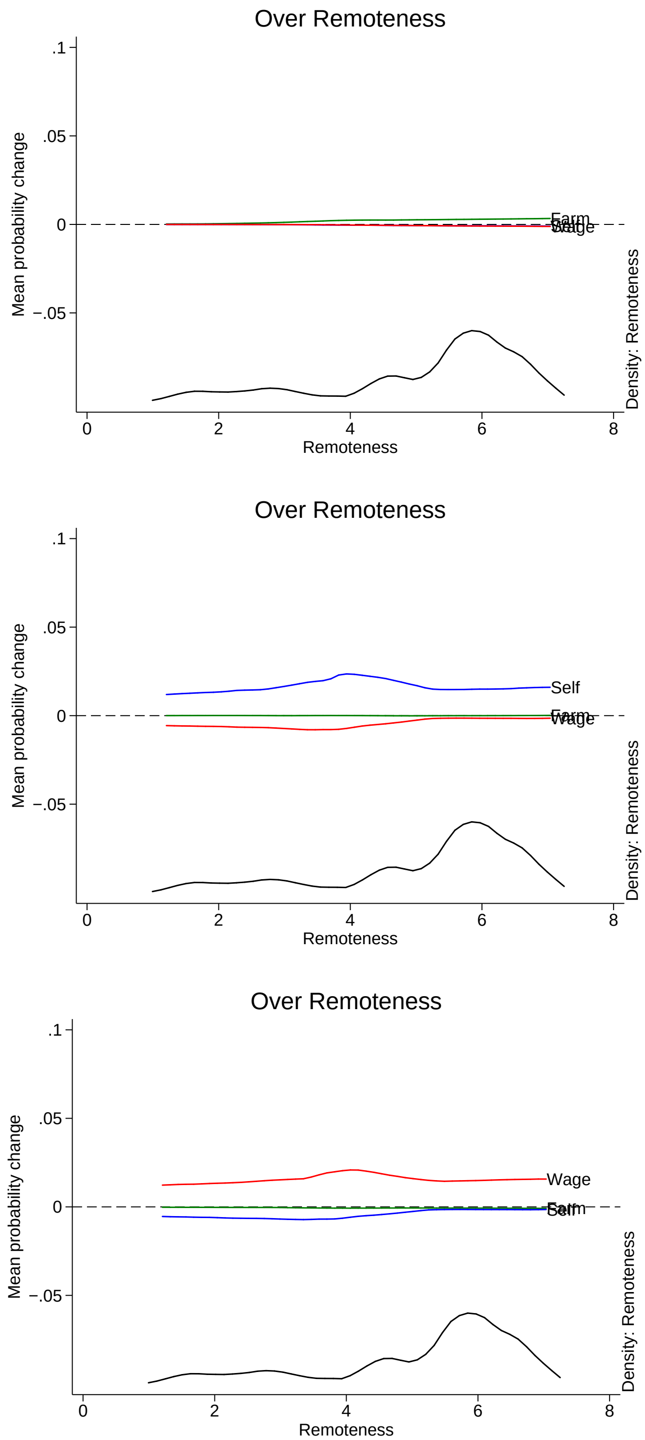


Figure 3.10: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over remoteness.

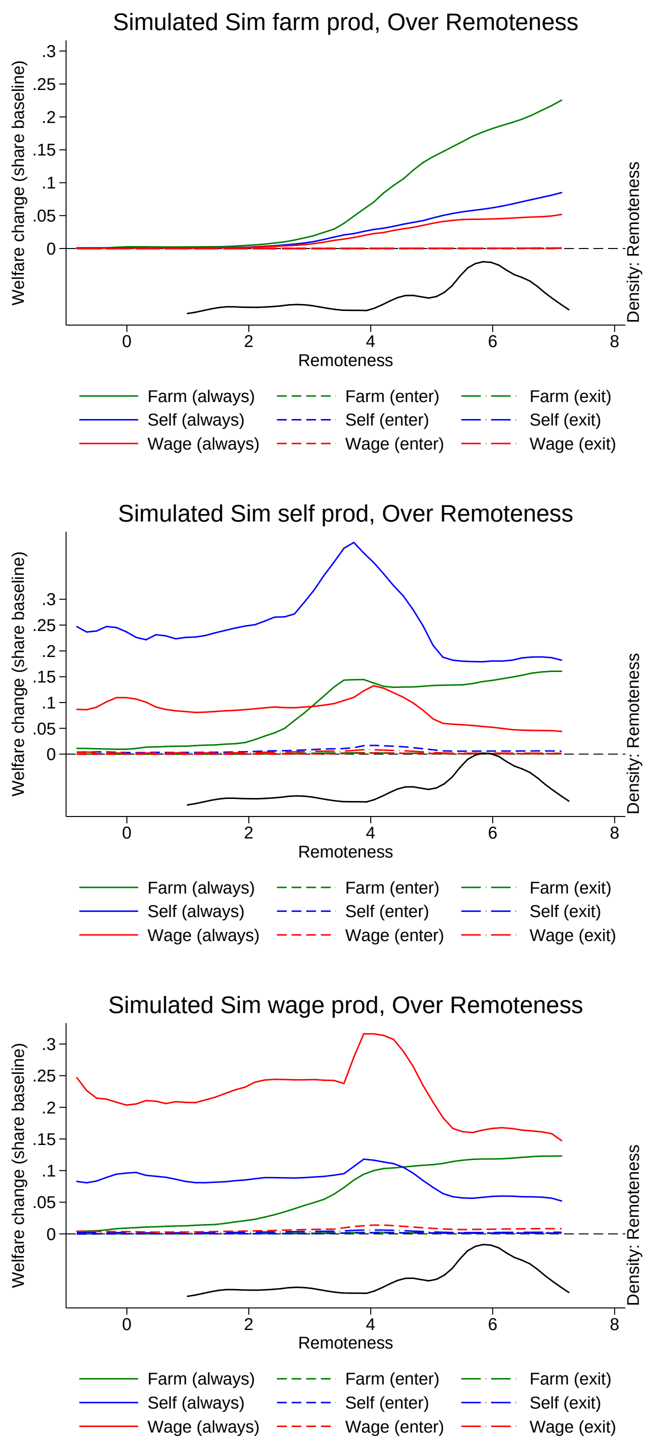


Figure 3.11: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on population density. The density of population density is shown underneath each regression.

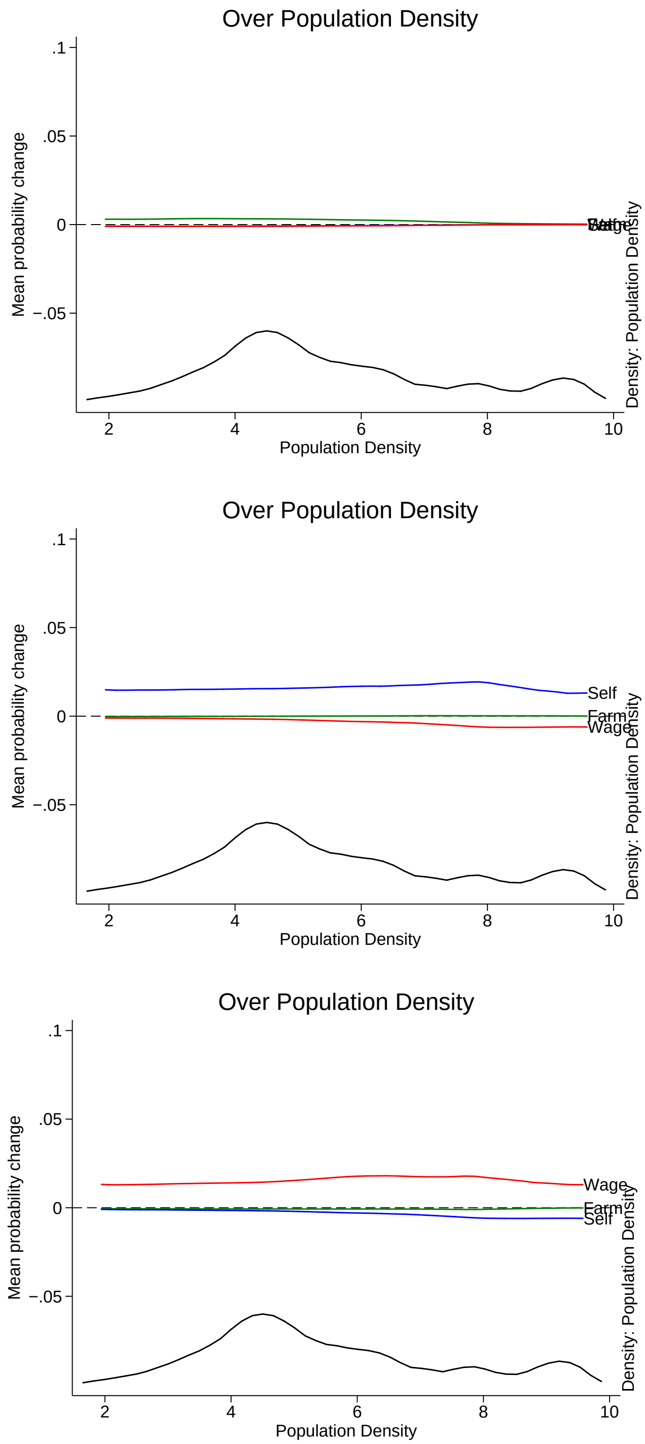


Figure 3.12: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over population density.

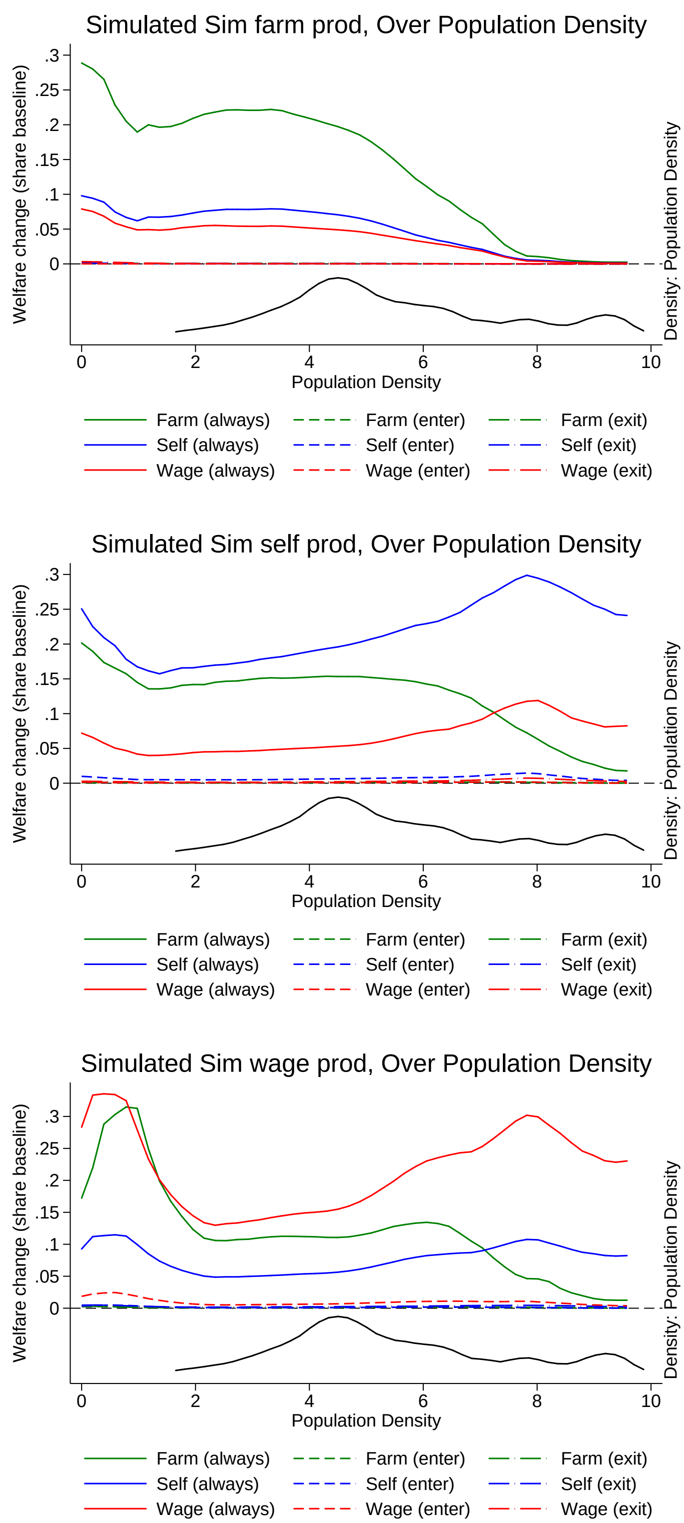


Figure 3.13: Simulated effect of doubling farm (top panel), self employment (middle panel), and wage labor (bottom panel) productivity on the expected change in probability of participating in farming, self employment, and wage labor conditional on agroclimatic potential. The density of agroclimatic potential is shown underneath each regression.

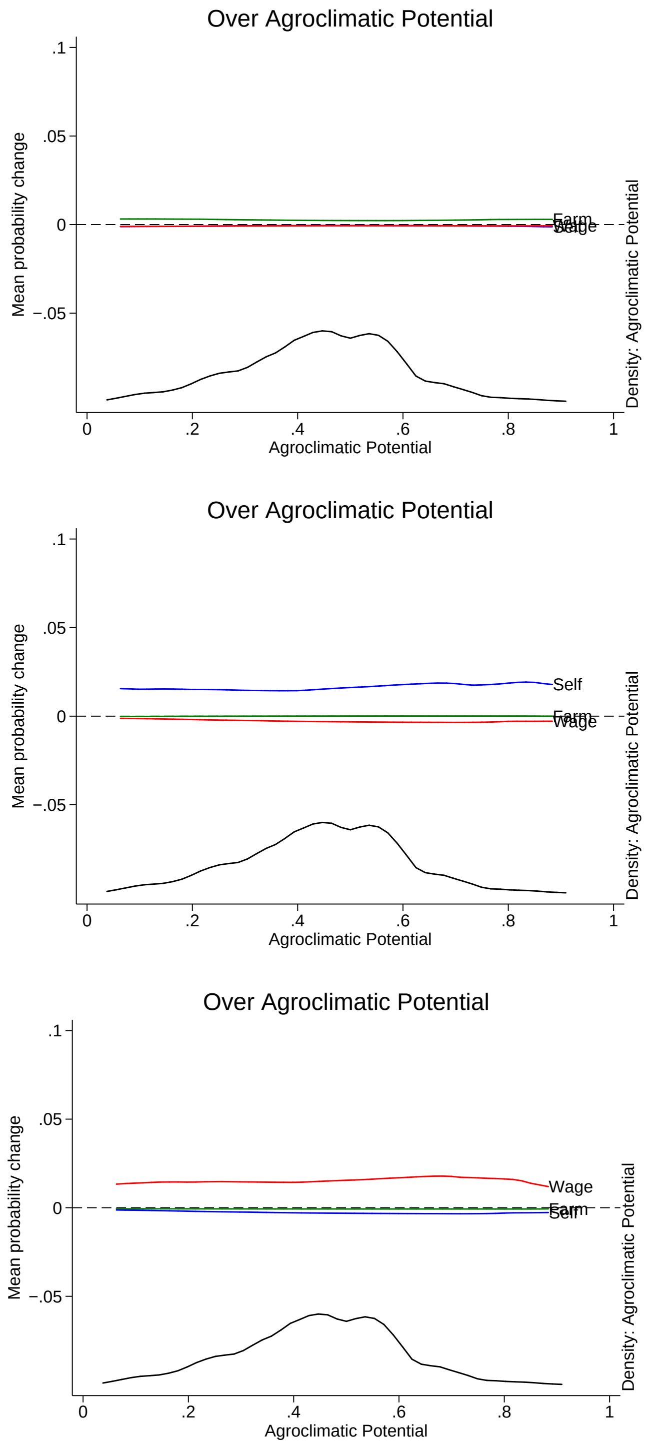
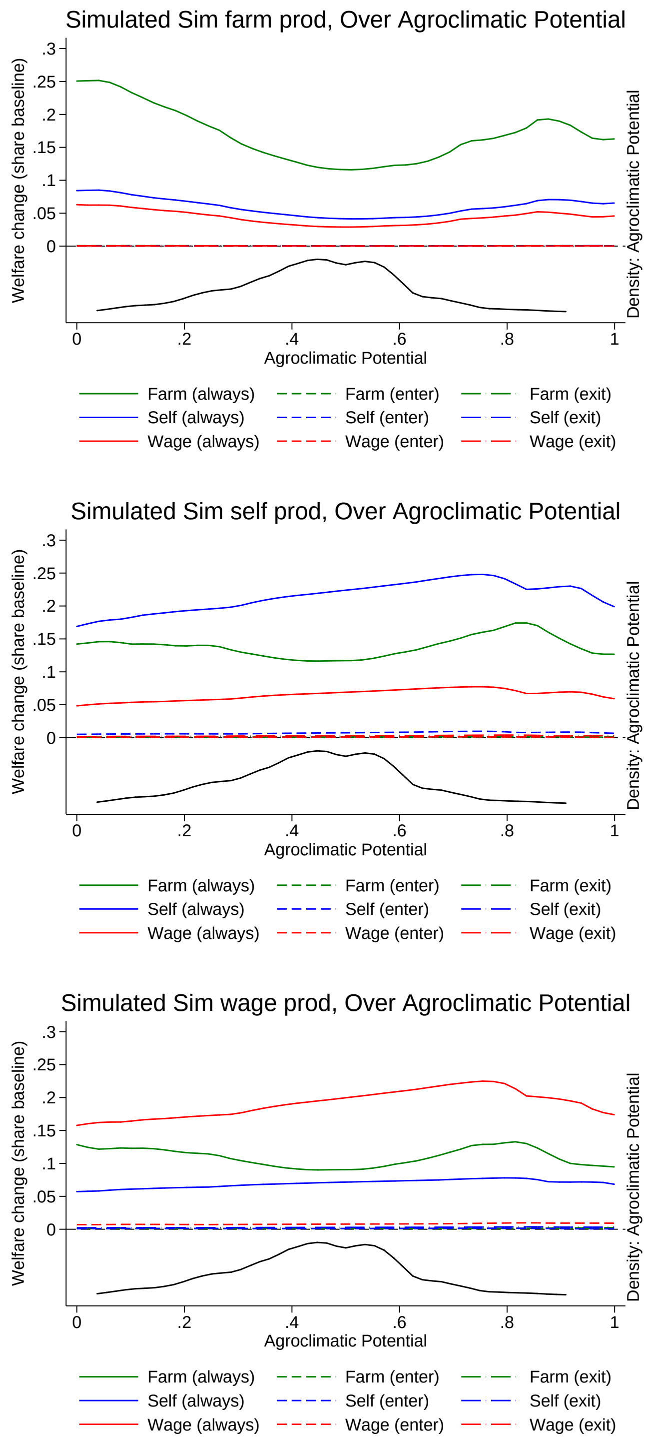


Figure 3.14: Expected utility gains for farm productivity simulation (top panel), self employment productivity simulation (middle panel), and wage employment simulation (bottom panel) over agroclimatic potential.



CHAPTER 4

MODELING PROFITABILITY OF AGRICULTURAL INTENSIFICATION INVESTMENTS WITH UNCERTAINTY AND SPATIAL HETEROGENEITY: DECISION SUPPORT FOR FERTILIZER TARGETING IN ETHIOPIA

(with Julianne Quinn, Andrew Simons, Leslie Verteramo, and Joshua Woodard)

4.1 Overview

According to the World Development Report, three out of every four poor people in developing countries live in rural areas, and most of these people depend on agriculture for their livelihoods (World Bank, 2008). In recognition of agriculture's importance within Sub-Saharan Africa, the region in which poverty remains most concentrated, African governments, NGOs, and many others have invested heavily in African agricultural development in recent decades. In 2003, several African heads of state committed, under the Maputo Declaration, to invest 10% of their countrys national budgets in agriculture. Within the broad agenda of agricultural development in Sub-Saharan Africa, soil health interventions have played a prominent role. They were featured in the Comprehensive Africa Agriculture Development Program (CAADP), a platform through which national governments design agricultural sector strategies. To date, ten governments have implemented major input subsidy programs, with costs on the order of $1 billion per year (Jayne and Rashid, 2013). Given the large expenditures associated with soil health programs, and growing concerns about cost effectiveness, many governments are now looking for ways to refine targeting and improve effectiveness of soil health interventions (Benin and Yu, 2013; Jayne and Rashid, 2013). Both uncertainty and spatial heterogeneity in the returns to agricultural investments complicate prioritization between investment categories and within programs (Jayne and Rashid, 2013).

Returns to agricultural investment are site specific, determined by soils, elevation, slope, prevailing climatic conditions, and the history of management at the site. Together, these agro-climatic features determine how crops grow, and how they respond physically to different management practices including applicatoin of inrganic fertilizer. Prices also vary from site to site, with location and infrastructure quality determining the costs of procuring inputs at the farm level. Output prices will be determined by local market conditions, which relate to the balance of supply and demand, storage infrastructure, and transportation costs to other markets (Benin and Yu, 2013).

Furthermore, agricultural investments are inherently uncertain, due to the time lag that separates investments and the realization of their returns. Most production in Sub-Saharan Africa is under rainfed conditions where the water needs of crops are met solely by in-season rainfall. Crops need water to synthesize carbohydrates through photosynthesis, and maize can be particularly vulnerable to water stress during specific stages of the growth cycle (Osgood et al., 2007). Yet farmers must make input purchases at the time of planting, before climatic conditions and output prices are known, forcing them to gamble on whether or not inputs will be profitable (Chavas and Holt, 1996).

Farmers also face uncertainty in how crops will respond to different inputs. This uncertainty, referred to as parametric uncertainty, complicates prediction of returns to agricultural investment. In a production function, parameters describe the relationship between inputs and outputs. A parameter value should not vary within the boundaries of the system that is being modeled. Challenges in estimating parameters can arise from poor model specification, problems with the data or methods used to estimate the parameters, or uncertainty in the model outcome (Walker et al., 2003). Even with a very cleanly identified parameter, one might worry about its validity in a slightly different setting or with a slightly different model specification. This paper will focus on the response of agricultural production to fertilizer application. Parameters estimating this response vary greatly, even within a country and for a given crop, due to interactions between growing conditions and the fertilizer response (Yanggen et al., 1998).

Currently, there is rapid growth in the availability of spatially explicit data in Sub-Saharan Africa that is relevant to an improved understanding of agricultural production relationships and price transmission. Only moderate inroads have been made in integrating such data for analytical purposes to inform large-scale, operational decisions and policies. Meanwhile, the underlying distributions for key climate variables are evolving with climate change. Incorporating changing climate variables into these operational decisions in order to support climate-smart agriculture poses additional challenges (Burke, Lobell, and Guarino, 2009).

To support investment planning and priority-setting in the face of risk and uncertainty in the developing country context, we create an *ex ante*, spatially explicit, profitability assessment tool. This allows users to visualize the probability of achieving a user-defined profitability objective given stochastic realizations of climate conditions and heterogeneous growing conditions.

We develop this decision tool in Ethiopia, a country that is highly committed to investing in agricultural growth and transformation, and one that is also characterized by high degrees of spatial heterogeneity, rainfall risk, and price risk Tadesse et al. (2006). Pro-poor agricultural growth is a pillar of the governments growth and transformation plan (Ethiopian Ministy of Finance and Economic Development, 2010). The growth and transformation plan prioritizes soil health improvement, with the Ethiopian government recently completing an intensive national soil mapping effort to generate sitespecific soil resource and fertility information with the purpose of improving soil management recommendations and decisions (Bomba, 2016). Having invested 14% of its budget in agriculture since 2003, the Ethiopian government continues to pursue a market-driven approach to smallholder productivity growth, one in which the government identifies and addresses bottlenecks that stand in the way of private sector investment. The government is also interested in promoting climate-smart agricultural growth, by adapting to a changing climate and building a system that is resilient to climate change (Bomba, 2016).

Soil health improvement has been a key component of the Ethiopian governments efforts. They have recently undertaken an intensive national soil mapping effort to generate sitespecific soil resource and fertility information with the purpose of improving soil management recommendations and decisions. Through a new fertilizer blending program, the government is investing in infrastructure and distribution systems to produce and deliver tailored fertilizer blends appropriate for local growing conditions throughout the country (Bomba, 2016). In the past, the government has made nationwide blanket recommendations that farmers apply 100 kg/ha each of urea and diammonium phosphate (DAP), a recipe that very few farmers follow according to recent national survey data (Sheahan and Barrett, 2014).

In this paper, we highlight the extent to which fertilizer profitability varies between different soil and climate conditions and explore the implications for decision makers who are designing and targeting soil health interventions. We initially develop this concept for nitrogen management on maize in Ethiopia, though the analysis can be expanded to other crops, nutrients, and management practices.

We develop a decision platform that integrates high resolution weather data, newly collected soil data, and extensive agronomic trial data to estimate the *ex ante* conditional yield response to inorganic fertilizer treatment in maize. Then, we use market price data to present a site specific, *ex ante* profitability distribution for fertilizer use. The decision platform allows users to visualize fertilizer profitability nationwide under different user-defined assumptions, such as for international fertilizer and crop prices, the desired profitability threshold, and risk tolerance.

We find that *ex ante* profitability assessments, which explicitly incorporate probability-oriented and spatially explicit factors, lead to different profitability conditions compared to the predominate approach to profitability assessment taking regionally estimated responses to fertilizer (derived experimentally or through observation) and valuing the costs and benefits according to recent prices. We discuss and explore relevant considerations for policy and decision-making.

4.2 Modeling Fertilizer Returns

The profitability of an agricultural technology is a key determinant of its adoption and use by farmers (Feder, Just, and Zilberman, 1985). Holding all else equal, it is unreasonable to expect farmers to adopt a technology if the value of the increased output generated is less than the cost of the technology. The Value-to-Cost Ratio (VCR) is a single measure that incorporates the private benefits and costs that a farmer faces when using an input or technology (Equation 1). Here, the change in output for crop y is depicted by Δy. The input, in this case, is fertilizer applied in quantity q (f_q). The output price is p^y, and the fertilizer price is p^f.

$$VCR_y(f_{dq}) = \frac{\delta y \cdot p^y}{dq \cdot p^f} \tag{4.1}$$

While profitability is a very powerful concept in economic analysis, VCR is often used quite bluntly. Profitability is typically assessed using aggregate area statistics or a few data points from model farms. This is problematic for several reasons. Profitability of input use varies systematically from site to site, with geographical features such as slope, aspect and elevation that influence temperature and precipitation; with underlying soil properties; with the farmers transport costs, which affect farm gate prices for inputs and outputs; with the farmers transport costs, which affect farm gate prices for inputs and outputs; and over time, as management practices affect productive capacity. As a result of site-specific heterogeneity in the determinants of input use profitability, farmers who are located in remote and agriculturally less favorable locations will systematically face lower profitability than the regional average, while farmers in less remote and more favorable locations will face higher than average profitability. Decision-makers can refine their predictions of profitability by accounting for these geographical determinants of profitability. Because the determinants of input use profitability, i.e., market access and agricultural favorability, are often correlated with socioeconomic variables such as poverty (Chamberlin and Schmidt, 2012; Harou et al., 2013; Dercon and Christiaensen, 2011), improved intervention targeting has both efficiency and distributional implications.

A second problem with profitability assessment is that it does not explicitly account for the uncertainty that farmers face when determining, in expectation, whether adopting a technology or input is likely to be profitable. It is important to recognize that farmers make their decisions in the presence of uncertainty regarding the marginal value of output that will result from use of a technology or input. Uncertainty can be found in crop responses depending on stochastic climate realizations and also in market fluctuations of input and output prices, since farmers generally do not know the market price they will receive for their crop when inputs are purchased. According to economic theory, farmers will optimally use less fertilizer when the output distribution faced is more variable (Anderson and Hardaker, 2003). When point estimates of profitability are proposed, these necessarily imply known crop response to the input and known output prices (Spielman, Kelemwork, and Alemu, 2011; Morris et al., 2007). However, assessing profitability in this way is only helpful for determining *ex post* whether using an input was profitable. Since farmers do not have the benefit of 20/20 hindsight when they must decide whether or how much to invest in an input, using a profitability measure that does not account for this uncertainty could lead one to conclude that farmers are under-using a technology when they are not.

Researchers have recognized the limitations of VCR as a metric to predict input use profitability and input demand by farmers. Often, they account for these shortcomings of the VCR measure by adjusting the target profitability threshold. That is, even though a VCR need only exceed 1 for the technology to be profitable, researchers look for a VCR greater than 2 to be confident that the benefits of using an input outweigh the costs (Morris et al., 2007; CIMMYT Economics Program, 1988). The assumption is that if the VCR is big enough on average, then profitability is robust, even though many farmers will have lower than average VCR, and risk may also play an important role in farmers decisions. Our concept of robustness, by contrast, relates not only to how profitability, on average, compares to a threshold but also how uncertain that profitability outcome is. Given the same expected profitability, one would expect farmers to perceive a technology to be more robust if profitability falls below the threshold once every ten years compared to once every three years.

In Equation 4.2, we introduce a site- and year- specific value-to-cost ratio (VCR) for a representative farmer in location i at time t applying q quantity of fertilizer (f) at unit price p_i^f to crop y:

$$VCR_{it}^{y}(f_q, \bar{\theta}, X_i, \omega_{it}) = [y(f_q, \bar{\theta}, X_i, \omega_{it}) - y(f_0, \bar{\theta}, X_i, \omega_{it})] \cdot \frac{p_{it}^{y}}{q \cdot p_t^{f}} \tag{4.2}$$

$\bar{\theta}$ refers to a vector of inputs and technologies used (e.g., seeds, non-fertilizer inputs), X_i is a vector of location conditions (soils, elevation, slope, etc.), and ω_{it} is a vector of climate variables in location i and period t. The VCR contains the expression for the average agronomic efficiency per kg of fertilizer applied compared with an application of zero fertilizer, $(y(f_q) - y(f_0))/q$.

In the *ex ante* framework, VCR is a random variable with probability distribution $f(VCR)$. The intent is not to study farmer behavior under risk, but just to understand the distributional aspects of returns to fertilizer within a given location. Input use decisions, both regarding fertilizer use f_q and other inputs $(\bar{\theta})$, are determined exogenously in this framework, meaning that the fertilizer response function applies to a model farmer who is hypothetically assigned to either use or non-use of fertilizer, and the fertilizer use decision is not modeled. Within each location, VCR_i is a function of a vector of climate variables, ω_i. Prices (p_i^y) are considered orthoganal to climate variables, which is consistent with grain price behavior in a small, open economy.

For a given profitability threshold T, we will evaluate robustness according to the probability that VCR is expected to exceed the threshold. This can be derived from the cumulative distribution function for VCR in location i:

$$Pr(VCR_i > T) = 1 - F(VCR_i(T)) \tag{4.3}$$

Because ω is a vector not a scalar, it makes sense to use a Monte Carlo approach to derive the CDF of VCR_i in each location, using boostrap sampling from a historic precipitation and temperature dataset (ω) that spans all of the sites. We describe the process by which the synthetic weather data are generated in the next section.

4.2.1 Understanding fertilizer response

In order to estimate the CDF of VCR in a given location, we must first estimate the parameters of a representative farmers production function, $y(\cdot)$. Of particular interest is the extent to which the agronomic response to fertilizer, $y(f_q) - y(f_0)$, varies with site characteristics (X), with climate realizations (ω), and with the other technologies and management practices employed ($\bar{\theta}$). There is lots of agronomic evidence, from East Africa and elsewhere, that the agronomic response to fertilizer depends on rainfall and temperature, soil conditions, technologies used, and other soil health practices (Yanggen et al., 1998; Vanlauwe and Giller, 2006).

Soil organic matter (SOM) can influence soil structure, moisture retention, and nutrient retention in soil, which is important because applied nitrogen leaches readily through the soil profile, becoming unavailable to crops (Magdoff and Van Es, 2000). Marenya and Barrett (2009) show that the agronomic yield response to fertilizer varies with SOM in Western Kenya. Initial soil mapping efforts in Ethiopia suggest low SOM levels in the highlands, especially. Soil pH also influences nutrient retention and availability to plants, with fertilizer-SOM and fertilizer-mineral interactions typically weakened as soils become more acidic. Soil micronutrients generally become more soluble in acidic soils, which can increase their availability to crops Sarkar and Wynjones (1982). About 41% of Ethiopian soils are classified as acidic, with the majority of the acidic soils found in the western region of Ethiopia (Schlede, 1989).

This spatial heterogeneity in precipitation is important to consider when assessing the returns to fertilizer in a given location. The majority of agriculture in Ethiopia is rain-fed, and empirical evidence suggests that rainfall is the common yield-limiting factor among all major cereals (Bewket, 2009). Haefele et al. (2006) show that fertilizer response decreases with increasing water stress during the growing season. To the extent that crop responses to soil health interventions are determined by rainfall levels, rainfall conditions in a single year will be a strong determinant of the profitability in that year of soil health interventions such as fertilizer application.

Temperatures also vary in space and are important determinants of crop growth. Lobell et al. (2011) show that there is a nonlinear response between temperature (growing degree days) and yields in African maize. Their results also suggest that nitrogen application can help mediate the effects of heat stress. Using side by side comparison of fertilizer treated and non-treated on-farm experimental plots across a large sample of Malawi farms over multiple growing seasons, Harou et al. (2013) find that fertilizer response varies with temperature and rainfall. Uyovbisere and Lombim (1991) repeat agronomic trials over multiple years and also find that fertilizer responses vary with rainfall and temperature.

4.2.2 Estimating parameters of a fertilizer response function

Because of the important interactions between fertilizer response, site-specific characteristics, management practices, and climate, it is crucial to estimate the parameters of a production function in order to predict fertilizer response. The function should be specified with sufficient flexibility to allow for these interactions that have been observed empirically. Particularly, the production function should not be additively separable with respect to fertilizer use and the other variables climate, location characteristics, and other technologies. Equation 4.4 shows a basic approach to modeling VCR using the modeled fertilizer response and prices.

$$VCR^{y}_{it}(f_q, p^{f}_{i}, \bar{\theta}) = [dy(f_q, f_0, \bar{\theta}, X_i, \omega_{it}] * \frac{p^{y}_{it}}{q * p^{f}_{t}} \tag{4.4}$$

There are four typical approaches to estimating the parameters of a fertilizer response function. First, agronomic trials are typically used to compare two plots with different fertilizer doses while holding all other variables constant. Some agronomic trials address a few additional parameters, such as crop variety or other crop management practice, through a factorial design. Measures of agronomic efficiency of fertilizer used on maize in East Africa exhibit a large spread. High responses are typically around 25 kg maize per kg nitrogen, while low responses are around 5 kg maize per kg nitrogen (Heisey and Mwangi, 1997). Agronomic trials are typically characterized by a very small sample sizes without a lot of variation in climate and soil conditions, and a limited number of treatment doses from which a crop response curve can be derived. It can be difficult to capture the interactions between fertilizer use and other variables using an agronomic trial dataset, either because the other variables are not recorded or because they do not vary across the observations. Though rainfall and temperature may not be recorded as part of an experimental study, they could be recovered from historical data using the location of the trial and historical climate data. However, one may not be able to confirm that supplemental water was not added to the crops under the trial. Furthermore, one cannot fully identify the contributions of different interactions between fertilizer and other variables to the local average treatment effect. The fertilizer response parameter, therefore, has limited external validity. And agronomic trial datasets are not well-suited for estimating other parameters of the production function.

Another concern about using agronomic trial data to estimate production function parameters is that they are often not conducted in locations that are representative of farmers fields (Nelson, Voss, and Pesek, 1985). Often, yield responses are higher in experimental stations than on farmers fields (Yanggen et al., 1998). Higher use of complementary inputs and more intensive weeding could bias fertilizer response upwards in experiment stations relative to farmers fields (Heisey and Mwangi, 1997). Studies have found nutrient responses often tend to be larger in depleted soils, where nutrients are limiting. If nutrients are more likely to be limiting on farmers fields than on experiment stations, then the fertilizer response at experiment stations will be biased downwards relative to farmers fields.

Model farm trials are a second source of data from which to estimate fertilizer response parameters. These trials typically involve side-by-side comparisons between fertilizer treatments on different farmers fields. Usually, but not always, the crop production is managed by farmers rather than scientists supervising the studies. Fertilizer practices are randomly assigned at the farm level, and typically the farms cover varying soil and climate conditions, allowing one to estimate key parameters of the fertilizer response function. This was the approach followed by Harou et al. (2013). The main concern with these datasets is that model farmers do not represent the farming population as a whole. Typically, they are better educated, more closely tied in with the extension system, and may differ in other characteristics. Fertilizer response parameters generated from model farm trials tend to be smaller in magnitude than those from experimental trials (Yanggen et al., 1998).

Observational farm surveys comprise a third source of data that can be used to estimate crop production function parameters. The main challenge with this approach is that fertilizer use is not randomly assigned, and is likely correlated with unobservable characteristics such as expected returns to fertilizer use and farmer ability, which then can bias the parameter estimates. Fertilizer response measured from observational data tends to be the smallest of the three categories (Yanggen et al., 1998). A recent comparison of nitrogen use efficiency measures derived from surveys with those from agronomic trials in Malawi suggests that farmer management practices, such as weeding, crop rotation, and timing and intensity of inorganic fertilizer application, can explain why nitrogen responses are lower on farmer fields than in research stations (Snapp et al., 2014).

The last approach to parameter estimation is through the use of a fully mechanistic crop growth model. These models are generally highly sensitive to data inputs (e.g., timing of fertilizer application, daily rainfall and solar radiation). And they are typically not calibrated to local conditions, which would require additional experimental or observational data anyway.

Given the strengths and weaknesses of the above approaches, a promising alternative is to assemble a meta-experimental dataset, pairing individual trial data points with their respective rainfall, locational, and management practice control variables. Then, the dataset as a whole can be used to estimate production function parameters, following Lobell et al. (2011). This allows for estimation of the interactions between climate and soil conditions and fertilizer response, while ensuring that fertilizer treatment is experimentally assigned, thus eleminating bias from selection into fertilizer use.

4.3 Data

We estimate the parameters of the crop response to fertilizer using the dataset of multiple maize trials across Eastern and Southern Sub-Saharan Africa compiled by Lobell et al. (2011). The dataset includes trials managed by the International Maize and Wheat Research Institute (CIMMYT), national agricultural research institutes, and private seed companies. The trials span 9 different years (1999-2007), 18 different countries, and 9 different agro-ecological zones. In this analysis, we initially focus on maize production because more farmers grow maize (in the Ethiopian Agricultural Sample Survey) than any other crop.[1]

Each crop trial was conducted to test variety performance, though the dataset was compiled in order to study crop response to water and temperature stress (Bänziger et al., 2006; Lobell et al., 2011). Fertilizer response was never an intended use of the data, although many of the varieties were tested under a low-nitrogen management regime, in which crops were planted on fields that were depleted of nitrogen due to continuous cropping of maize over previous seasons, removing all stover after harvest, and not applying any organic or inorganic fertilizer. In optimal-management trials, the recommended amounts of nitrogen fertilizer were added. All other crop management practes were held constant between low-nitrogen and optimal-management trials.

Using the locations of each experiment site, we matched the crop trial data with climate data, including monthly total precipitation and monthly average

[1]This approach can be expanded to other crops with additional trial data.

temperatures over the growing season. Weather data are obtained from the United States National Oceanic and Atmospheric Administration (NOAA). We use mean monthly temperature estimates at 0.5 degree resolution from NOAAs National Climatic Data Center (NCDC) combined Global Historical Climatology Network (GHCN) version 2 and Climate Anomaly Monitoring System (CAMS) analysis.[2] We use daily precipitation estimates at 0.1-degree resolution from NOAAs Climate Prediction Center (CPC) Africa Rainfall Estimate Climatology Network version 2 (ARC2) dataset.[3] In the crop growth period, which we define as five months after planting (Bänziger et al., 2006), we calculate monthly total precipitation and average temperature for each site. The third month generally coincides with flowering and silking, a period that is considered especially sensitive to water and temperature stress. For temperature data we use the average period temperature, while for precipitation we use the accumulated precipitation for the same period.

We match the trial sites with soil data from the Africa Soil Information Service (AfSIS).[4] The 250 meter resolution soil data include estimates of several soil characteristics at different layers, such as soil cation exchange capacity, pH, texture, and water retention capacity. Finally, we match the trial data with Agro-ecological zone (AEZ) classificatoins from GAEZs and are downloaded through Harvest Choice.[5]

Table 4.1 shows mean descriptives in low-nitrogen and optimal-nitrogen sites, along with the results of a T-test for comparison of means between these sites. Yields in the low nitrogen sites are 1.99 t/ha, which is about half of yields

[2]http://www.esrl.noaa.gov/psd/data/gridded/data.ghcncams.html
[3]ftp://ftp.cpc.ncep.noaa.gov/fews/fewsdata/africa/rfe2
[4]http://www.isric.org/content/african-soilgrids-250m-geotiffs
[5]http://harvestchoice.org/maps/agro-ecological-zones-sub-saharan-africa.

in optimally managed sites, which average 3.89 t/ha. The low-nitrogen trial sites differ from the optimal fertilizer sites in climate and soil conditions as well. These key differences across sites arise from the fact that low-nitrogen trials tended to be concentrated in lowland, sub-humid areas. Even though the differences are statistically significant for all of the climate and soil variables, they are not especially large in magnitude. Because nitrogen treatment was experimentally assigned, rather than selected endogenously by farmers, a clean identification of the effect of nitrogen on crop growth is ensured.

4.4 Results

We estimate a flexible, quadratic random effects yield model with GLS, as depicted in Equation 4.5. Maize yield (in metric tonnes of grain per hectare) is the dependent variable, and the model includes AEZ fixed effects. Optimal fertilizer management sites are assigned a fertilizer treatment dummy of one, while low-nitrogen management sites were assigned a fertilizer treatment dummy of zero. The design does not allow for estimating a continuous fertilizer dosing effect on crop growth. It is, however, appropriate for estimating the binary impacts of adopting fertilizer at the level recommended by agronomists. The fertilizer treatment dummy is interacted with all of the other yield function variables. We do not include site or year fixed effects because the purpose of generating parameter estimates is to predict fertilizer response outside of the crop trial sample. Standard errors are clustered at the site-year level following Lobell et al. (2011). The estimation sample is restricted to trial site-years that fall within the temperature and precipitation range observed in Ethiopia.

Table 4.1: Summary statistics of model variables by fertilizer management strategy.

	Low Nitrogen mean/sd	Optimal Fertilizer mean/sd	T-test b/t
Yield (t/ha)	1.99	3.89	-1.90$***$
	(1.30)	(2.54)	(-59.74)
Temp months 1-2 (mean, $^\circ$ C)	23.38	22.41	0.97$***$
	(2.04)	(2.69)	(21.56)
Temp month 3 (mean, $^\circ$ C)	22.67	22.08	0.59$***$
	(1.91)	(2.59)	(13.94)
Temp months 4-5 (mean, $^\circ$ C)	21.04	20.79	0.25$***$
	(1.71)	(2.67)	(6.29)
Precip months 1-2 (tot, mm)	231.43	252.30	-20.88$***$
	(124.84)	(126.94)	(-7.88)
Precip month 3 (tot, mm)	134.56	102.05	32.52$***$
	(96.95)	(92.08)	(15.94)
Precip months 4-5 (tot, mm)	101.10	83.93	17.18$***$
	(118.72)	(89.72)	(7.02)
Soil cation exchange capacity (centimol charge per kg soil)	12.93	13.71	-0.78$***$
	(6.91)	(8.17)	(-5.20)
Soil pH (pH determined in soil/water mixture)	5.92	6.06	-0.13$***$
	(0.43)	(0.44)	(-14.61)
Soil clay (share by volume)	0.37	0.28	0.09$***$
	(0.21)	(0.10)	(20.93)
Soil silt (share by volume)	0.17	0.17	-0.01$***$
	(0.05)	(0.06)	(-6.32)
Poor drainage (dummy)	0.14	0.17	-0.04$***$
	(0.34)	(0.38)	(-4.80)
Observations	2543	18226	20769

$$Y_{it} = \beta_0 + \sum_j \beta_j X_{jit} + \sum_j \sum_k \frac{1}{2} \beta_{jk} X_{jit} X_{kit} + F_{it} * \left[\sum_j \beta_{Fj} X_{jit} \right] + \sum_{AEZ} \beta_{AEZ} \delta_{AEZi} + \varepsilon_{it} \quad (4.5)$$

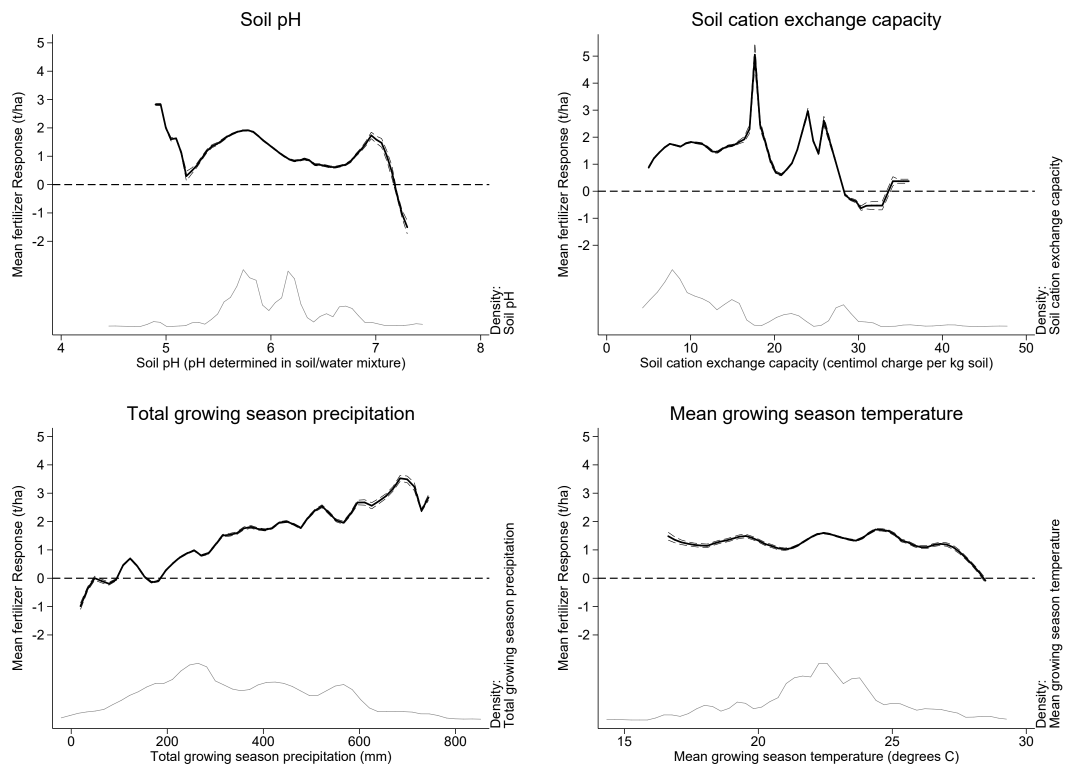
Because there are a large number of parameters given all of the interactions specified, the elasticities at the means of the dataset are shown in Table 4.2. The elasticities and their standard errors are shown at the mean of the full estimation sample, and at the mean of the subsample of trial sites within Ethiopia. Under optimal fertilizer application, temperature has a negative effect on yields during the first two periods of crop growth and a positive effect during the third period. Precipitation during all three periods of growth has a slightly negative effect on yields at the mean of the data. And yields are decreasing in soil pH. The Ethiopian trial sites are similar to the sample as a whole, except yields are increasing in precipitation and soil silt share at the means of the Ethiopian subsample. When fertilizer is withheld, temperature has a positive effect in the second period and a negative effect in the third period. Precipitation has a positive effect in the first period, and soil cation exchange capacity has a positive effect.

Table 4.2: Average marginal effect of explanatory variables (soil, climate and location characteristics) on yield, with and without fertilizer use. The first and third columns depict mean effects for all trial sites, while the second and fourth columns depict mean effects for Ethiopian trial sites only.

	No Fert. All	No Fert. Eth	Optimal Fert. All	Optimal Fert. Eth
Temp months 1-2 (mean, $^\circ$ C)	-0.00349***	0.0292	-0.00102***	-0.0867
	(0.000845)	(0.0913)	(0.000250)	(0.0623)
Temp month 3 (mean, $^\circ$ C)	0.00286***	-0.678***	-0.00170***	-0.178**
	(0.000835)	(0.157)	(0.000227)	(0.0646)
Temp months 4-5 (mean, $^\circ$ C)	-0.00576***	0.624***	0.000644***	0.296***
	(0.000957)	(0.112)	(0.000182)	(0.0468)
Precip months 1-2 (tot, mm)	0.000880*	-0.0294**	-0.00177***	0.0220**
	(0.000375)	(0.0112)	(0.000165)	(0.00683)
Precip month 3 (tot, mm)	-0.00304***	0.0225***	-0.00297***	0.0297***
	(0.000456)	(0.00596)	(0.000245)	(0.00491)
Precip months 4-5 (tot, mm)	-0.000241***	0.0957***	-0.0000971***	0.0436**
	(0.0000391)	(0.0249)	(0.0000135)	(0.0149)
Soil cation exchange capacity (centimol charge per kg soil)	3.49e-09**	0.0500	-1.23e-10	-0.560
	(1.09e-09)	(0.482)	(7.04e-10)	(0.388)
Soil pH (pH determined in soil/water mixture)	-3.02e-10	-0.00404	-3.89e-10**	-0.0146
	(2.12e-10)	(0.0209)	(1.34e-10)	(0.0167)
Soil clay (share by volume)	-1.65e-10	-0.133	-7.53e-11	-0.0903
	(3.79e-10)	(0.218)	(2.46e-10)	(0.170)
Soil silt (share by volume)	-1.28e-08***	0.567	-3.22e-10	0.905**
	(3.68e-09)	(0.418)	(2.20e-09)	(0.326)
Observations	20,486	369	20,486	369

In order to better understand how fertilizer response varies under different growing conditions, we predicted the difference between an optimal-nitrogen crop and a low-nitrogen crop for each trial site, holding all climate and soil characteristics to the values observed in the dataset. The predicted fertilizer response is then graphed non-parametrically over climate and site characteristics (Figure 4.1). The densities of the climate and site characteristics are also depicted at the bottom of each graph. The graphs show that expected fertilizer response is variabel but decreasing, on average, over soil pH and cation exchange capacity. Expected fertilizer response is increasing in growing season precipitation and decreasing in temperature.

Figure 4.1: Mean predicted fertilizer yield response over soil pH (top left), soil cation exchange capacity (top right), total growing season precipitation (bottom left), and average growing season temperature (bottom right).



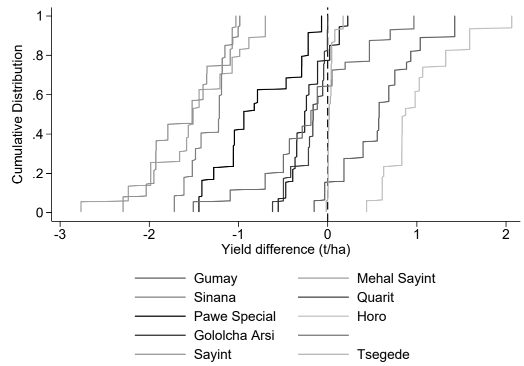
Using the production function parameters estimated above, we next simulate the angronomic response to fertilizer across Ethiopian geographies. We converted gridded soil and AEZ data into area-weighted averages for each *woreda* in which at least some maize is grown.[6] Using each *woreda*'s predominant AEZ, we predict maize planting month using the FAOs crop calendar.[7] We match this *woreda* level dataset with fourteen years of temperature and precipitation data using the protocol used to match the crop trial sites with temperature and precipitation data, averaged at the *woreda* level.

We generate a 200-year synthetic weather dataset at the *woreda* level by randomly sampling from the fourteen years of available temperature and precipitation data. Using this synthetic dataset, we examine the characteristics of fertilizer response for each *woreda* over stochastic climate conditions. The cumulative distributions of of fertilizer response are shown for ten randomly selected *woredas* in Figure 4.2. Some of the *woredas* are characterized by low responsiveness to fertilizer, while others are characterized by high responsiveness. And the *woredas* also differ in the variability of fertilizer response across climate iterations. The simulations show that, in some *woredas*, fertilizer response is always predicted to be positive, while in others, it is always predicted to be negative. For most *woredas*, fertilizer response is sometimes positive and sometimes negative.

[6] We use Ethiopian Agricultural Sample Survey data, available from 2007 and 2012, to screen for nominal area planted to maize in each Ethiopia *woreda*.

[7] http://www.fao.org/agriculture/seed/cropcalendar/welcome.do

Figure 4.2: Cumulative distribution of yield response to fertilizer over stochastic weather and precipitation conditions for ten different randomly selected *woredas*.



4.5 Decision Support

Next, we turn to analysis of the profitability of fertilizer use. We convert the predicted yield difference into a value cost ratio (VCR) measure using an assumed fertilizer price of 5.4 ETB per kg ($0.26 USD) (Rashid et al., 2012), and an assumed maize price of 5,000 ETB per MT ($250 USD). We can then analyze profitability, *ex ante*, according to properties of the distribution of the stochastic VCR variable (see Equation 4.6). For the purposes of this analysis, we assume that a farmer seeks at least a 20% return on the fertilizer investment (T=1.2) at least 70% of the time (P=0.7).

$$1 - F(VCR = T) > P \tag{4.6}$$

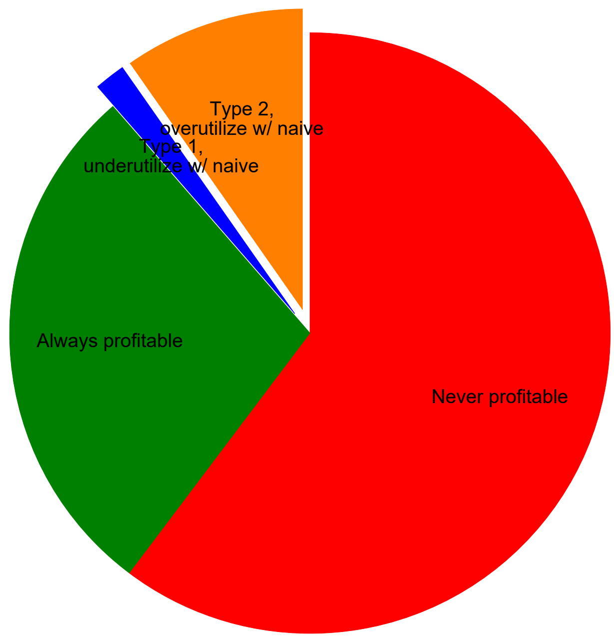
We can then characterize *woredas* by whether the robust profitability criteria specified in Equation 4.6 are met, and explore the implications for decision makers. For the sake of comparison, we define a "naive" profitability measure according to the two most recent climate realizations. This measure is analogous to an *ex post* measure of profitability as commonly applied in the literature. After constructing both naive and robust fertilizer profitabilities for each *woreda*, we then compare the two.

At the desired profitability incidence of 70%, we find that fertilizer use is not deemed profitable by either “naive” or “robust” criteria in the majority of *woredas* in which maize is grown. In about a fourth of the “woredas”, fertilizer use is profitable using both “naive” or “robust” criteria. In about 10% of the *woredas*, fertilizer use would be considered profitable according to “naive”, *ex post* criteria, but not according to robust, ex ante criteria. In these cases, one might over-predict the returns to fertilizer use if one does not fully consider stochastic weather realizations. In only 1.5% of the *woredas* is fertilizer use considered profitable according to “robust”, *ex ante* criteria but not “naive”, *ex post* criteria. In these cases, a farmer might under-predict the returns to fertilizer use without considering stochastic weather realizations. Because the very recent years tend to be better, on average than the full climate record, the Type 2 classification, whereby fertilizer profitability is over-estimated, is more common.

4.6 Conclusion

We have proposed a flexible approach to assisting decision makers in assessing the returns to soil health investments in the face of uncertainty and spatial heterogeneity. Predicted fertilizer response in an agronomic trial setting may not perfectly correspond with fertilizer responses that farmers will observe on their fields. However, it is nevertheless informative to explicitly examine the interactions between fertilizer response, climate realizations, and site character-

Figure 4.3: Cumulative distribution of yield response to fertilizer over stochastic weather and precipitation conditions for ten different *woredas*.



istics. It indicates that profitability is likely to be quite sensitive to the criteria by which decision-makers define profitability. It is important to better understand these criteria when calibrating decision support tools, and when using predicted profitability to understand fertilizer adoption behavior.

This approach, as a platform, can be strengthened as more data become available. Additional fertilizer response trials within Ethiopia would be especially useful, as would better understanding the specific fertilizer management practises that were used across the trial sites. It would be of great interest to better understand the intensive margin of fertilizer response in Ethiopia, that is, the yield gain as one varies the fertilizer application rate between zero and the total recommended amount. However, many more fertilizer dosing trials covering multiple sites and years would be required for this. The response to soil micronutrient ammendment is also interesting, but will be difficult to model with existing data.

One feasible expansion would involve conditioning the synthetic climate data draw on the ENSO signal that is available at the time of planting, in order to further differentiate the prediction of fertilizer profitability in El Nino and La Nina years, when climatic patterns tend to differ (Korecha and Barnston, 2007).

APPENDIX A

APPENDIX FIGURES

APPENDIX B

APPENDIX TABLES

Table B.1: Wage Labor variable construction notes

	ETHIOPIA 2013-14	MALAWI 2010-11	TANZANIA 2010-11	UGANDA 2010-11
Job tracking	Up to two jobs, Productive Safety Net Program (PSNP) labor, and casual labor	Up to two jobs and casual (*ganyu*) labor	Up to two jobs	Up to 4 activities including jobs, own farm and NFE (2 most important in past 7 days and up to 2 more if they are more important over last year)
Time use aggregate	For jobs: months last year; typical weeks / month worked; typical hours / week worked. For casual labor (and PSNP in Ethiopia): days / last year (hours per day assumed)		Months last year; typical weeks / month worked; hours last week	Months / year; weeks / typical month worked; hours last week
Cleaning (time)	Assume unreported time observation = 0. Truncate unreasonably large values (e.g., >16 hours per day or >7 days per week)			
Sector information	Observed for jobs. PSNP is classified as ag. Casual labor is classified as "unknonwn"	Observed for jobs. Casual (*ganyu*) labor is classified as ag.	Observed for both jobs	Observed for all activities (for 4th activity sector is observed but not whether it is a job, family farm, or NFE)
Wage rate	Earnings reported per pay period (for jobs) and annually (for casual and PSNP labor)	Earnings reported per pay period (jobs) and per day (casual labor)	Earnings reported per pay period	Earnings reported per pay period (lots of missing pay period observations)
Cleaning (wages)	Winsorize hourly wage rate (p = 0.01) and reconstruct annual returns with imputed wage			
Annual returns	For each job: cleaned wage rate * annual hours			

Table B.2: Farm enterprise variable construction notes (for households involved in crop production).

<table>
<tr><th></th><th>ETHIOPIA
2013-14</th><th>MALAWI
2010-11</th><th>TANZANIA
2010-11</th><th>UGANDA
2010-11</th></tr>
<tr><td>Seasons included</td><td>Only main (meher) season</td><td>Rainy and dry seasons</td><td>Rainy and dry seasons</td><td>First and second seasons</td></tr>
<tr><td>Farm tasks included</td><td>Post planting (land prep, planting, ridging, weeding, fertilizer application); harvest</td><td>Land prep / planting, pre-harvest; harvest</td><td>Land prep; weeding; post planting; harvest</td><td>All tasks lumped together</td></tr>
<tr><td>Participation (indiv)</td><td>Up to 6 hh members identified per task per plot/task combo</td><td>Up to 4 hh members per plot/task combo</td><td>Up to 6 individuals per plot/task combo</td><td>Up to 3 individuals per subplot</td></tr>
<tr><td>Time use aggregate (indiv)</td><td>For each indiv/plot/task: days / week (assume 7 hrs/day), weeks / season</td><td>For each indiv/plot/task: days / week (assume 7 hrs/day), weeks / season</td><td>For each indiv/plot/task: days per season (assume 7 hrs/day)</td><td>For each sub-plot: days per season (assume 7 hrs/day, assume equal labor division b/w individuals listed)</td></tr>
<tr><td>Cleaning (time)</td><td colspan="4">Truncate unreasonably large values (e.g., an individual works more days than the length of the season)</td></tr>
<tr><td>Hired labor</td><td>For each plot/task: male, female and child hired laborer person-days</td><td>For each plot: male, female and child hired laborer person-days for pre-harvest and harvest periods</td><td>For each plot/task: male, female and child hired laborer person-days</td><td>For each plot: male, female and child hired laborer person-days</td></tr>
<tr><td>Exchange labor</td><td>For each plot/task: male, female and child exchange laborer person-days</td><td>For each plot: male, female and child exchange laborer person-days</td><td>Not in survey</td><td>Not in survey</td></tr>
<tr><td rowspan="2">Farm labor inputs (hours)</td><td>Own farm labor + hired and exchange labor</td><td>Own farm labor + hired and exchange labor</td><td>Own farm labor + hired labor</td><td>Own farm labor + hired labor</td></tr>
<tr><td colspan="4">In all countries, days of household and hired labor are converted to hours (7/day). Labor inputs supplied by children (age 5-14) are assigned a weight of 0.5</td></tr>
<tr><td>Farm labor inputs (people)</td><td colspan="4">Number of family members with positive hours working on farm (no count of hired or exchange workers available). Children (age 5-14) are assigned a weight of 0.5</td></tr>
<tr><td>Annual returns</td><td colspan="4">Net farm returns are taken from RIGA (see section 2 of paper). Quantity of auto-consumed farm production is estimated using the consumption module of the survey. Net returns are truncated at zero.</td></tr>
</table>

Table B.3: Non-Farm enterprise (NFE) variable construction notes.

	ETHIOPIA 2013-14	MALAWI 2010-11	TANZANIA 2010-11	UGANDA 2010-11
Participation (indiv)	HH indicates operation of NFE; up to 5 HH members can be listed as workers	HH indicates operation of NFE; up to 4 HH members can be listed as workers	Indiv indicates whether he / she is involved in self employed activities	HH indicates operation of NFE; up to 5 HH members can be listed as workers
Time use aggregate (indiv)	Predicted for non-reporting households using NFE months in operation and median 7-day recall	For each HH member listed as NFE worker: months / last year; typical days / month worked; typical hours / day worked	Predicted for non-reporting households using NFE months in operation and median 7-day recall	Individuals report NFE employment in labor module (months / last year; typical weeks / month; hours / last week)
Cleaning (time)	Truncate unreasonably large values (e.g., >16 hours per person per day).			
Sector information	Industry code provided in NFE module	Industry code provided in NFE module	Industry code provided in labor module	Industry codes missing for 2/3 of NFEs. They are predicted using industry coding from labor module.
Hired labor	Number employees working for NFE (past 12 months)	Number of male, female and child employees (and hours worked) in a typical month of operation	Number employees working for NFE (past month)	Number employees working for NFE (past month)
Firm labor inputs	Can recover number of workers (own and hired), not hours worked	Can recover number of workers and hours worked	Can recover number of workers (own and hired), not hours worked	Can recover number of workers (own and hired), not hours worked
Annual returns	Net firm revenues (reported gross sales minus reported costs)	Reported net firm revenues	Net firm revenues (reported gross sales minus reported costs)	Net firm revenues (reported gross sales minus reported costs)
Cleaning (returns)	Reported revenue and cost variables are winsorized (p=.01). Net revenue variable is truncated at 0			

Table B.4: Summary statistics of regressors

VARIABLES	N	mean	sd	min	max
Transfers recvd (USD)	3,599	76.62	150.8	0	1,315
Household size	3,599	5.313	3.062	1	35
HH dependent share	3,599	0.324	0.255	0	1
Head's father attended school	3,599	0.469	0.499	0	1
Years educ, head	3,599	6.518	4.530	0	20
Urban	3,599	0.320	0.467	0	1
Peri-urban dummy	3,599	0.0547	0.227	0	1
Ave length of season (days, MOD12Q2)	3,599	178.2	23.11	131	234
Hrs travel to nrst town >500k (LSMS-ISA)	3,599	6.315	4.812	0	20.35
Network dist to nrst town >100k (km, LSMS-ISA)	3,599	106.9	99.62	0.240	546.4
Dist nrst major rd (km, LSMS-ISA)	3,599	17.05	23.32	0	135.4
People per square km, 2005 (ln, HC)	3,599	5.430	1.914	0	9.578
Number hh members 16-65	3,599	2.835	1.735	0	24
Age of head	3,599	46.49	15.63	18	105
Female head	3,599	0.252	0.434	0	1
Yrs educ, adults (ave)	3,599	6.461	3.969	0	20
Yield Potential low (cross-crop ind)	3,599	0.444	0.158	0	0.999
Yield Potential high (cross-crop ind)	3,599	0.529	0.179	0	1
Rate improved maize seed use (mean smth)	3,599	0.116	0.176	0	1
Cost hired lbr (med smth, USD/day)	3,599	2.786	2.379	0.194	10.12
Rate of tractor use (mean smth)	3,599	0.0286	0.0856	0	0.800
Land owned (ha, RIGA)	3,599	1.192	1.890	0	19.83
Mean precip wettest qrtr (mm, NOAA CPC)	3,599	579.9	172.3	231	1,440
Slope (pct, USGS)	3,599	4.886	4.848	0	46.60
Soil nutrient retention capacity (FAO)	3,599	1.496	0.956	0	7
Soil workability (FAO)	3,599	1.513	1.128	0	7
Share land irrigated (percent, FAO)	3,599	0.00566	0.0275	0	0.288
Cost/hired worker (med smth, USD)	3,599	791.3	851.6	2.049	4,973
Nighttime light ave coverage(DMSP F16)	3,599	683.4	1,372	0	4,764
Financial service available	3,599	0.450	0.498	0	1
Productive non-ag asset ind, fact 1	3,599	0.876	0.734	0	4.281
Productive ag-related asset ind, fact 1	3,599	0.0322	0.0433	0	0.341
Max educ in hh (yrs)	3,599	7.993	4.298	0	20
Nighttime light intensity(DMSP F16)	3,599	8.313	13.61	0	47.64
Returns/ag worker (med smth, USD)	3,599	449.9	487.6	14.94	2,135
Returns/ind worker (med smth, USD)	3,599	1,525	1,237	149.4	5,124
Returns/ser worker (med smth, USD)	3,599	1,974	2,251	25.62	17,079
Partpn in ag emplmt (med smth sh)	3,599	0.109	0.147	0	1
Partpn in ind emplmt (med smth sh)	3,599	0.0632	0.0934	0	0.667
Partpn in ser emplmt (med smth sh)	3,599	0.233	0.202	0	0.833

Table B.5: Comparison of farm profit coefficients with and without selection, Generalized Leontief specification. The first model includes only farm participants (no selection). The second and third columns present marginal effects from the second and first stages of a Heckman selection model, respectively.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	-171.761	101.024	135.842	105.371		
Peri-urban dummy (sq rt)	-324.229	786.995	-11.949	678.455		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-36.261	48.643	-87.897	53.434		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	10.518	7.465	14.805	8.933		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	2.188	13.251	15.000	18.173		
People per square km, 2005 (ln, HC) (sq rt)	-163.464	107.090	-204.729	122.090		
Number hh members 16-65 (sq rt)	438.735**	52.232	360.126**	64.025		
Age of head (sq rt)	20.156	19.963	43.922	25.068		
Female head (sq rt)	-88.432	47.877	-54.241	56.824		
Yrs educ, adults (ave) (sq rt)	37.378	27.615	66.084	35.784		
Yield Potential low (cross-crop ind) (sq rt)	293.084	367.483	-185.002	460.278		
Yield Potential high (cross-crop ind) (sq rt)	-689.259	371.876	-94.349	436.320		
Rate improved maize seed use (mean smth) (sq rt)	-50.670	148.237	-184.248	136.459		
Cost hired lbr (med smth, USD/day) (sq rt)	-167.123*	66.562	-121.834	67.761		
Rate of tractor use (mean smth) (sq rt)	747.074*	290.517	407.344	323.785		
Land owned (ha, RIGA) (sq rt)	440.026**	33.071	383.472**	40.049		
Mean precip wettest qrtr (mm, NOAA CPC) (sq rt)	14.367	10.947	30.411*	12.090		
Slope (pct, USGS) (sq rt)	23.909	37.129	59.790	43.512		
Soil nutrient retention capacity (FAO) (sq rt)	-148.596	105.878	-217.076	119.507		
Soil workability (FAO) (sq rt)	2.062	98.096	216.567	130.760		
Share land irrigated (percent, FAO) (sq rt)	1,987.293*	860.207	2,556.904**	899.619		
Transfers recvd (USD)					-0.001**	0.000
Household size					0.099**	0.010
HH dependent share					0.623**	0.106
Head's father attended school					-0.035	0.055
Years educ, head					-0.063**	0.006
Urban					-1.391**	0.055
Peri-urban dummy					-0.619**	0.103
Ave length of season (days, MOD12Q2)					-0.011**	0.001
N	2,407		3,599		3,599	
R2 (adj)	0.343		0.327			
Lambda			-390.54			
Sigma			800.66			
P value comparison test			0.00			
N (non-censored)			2407			

* $p < 0.05$; ** $p < 0.01$

Table B.6: Comparison of self employment profit coefficients with and without selection, Generalized Leontief specification. First stage Generalized Leontief estimation of enterprise profits. The first model includes only enterprise participants (no selection). The second and third columns present results from the second and first stages of a Heckman selection model, respectively. The marginal effects of profit function variables are shown.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	-403.709	392.752	-641.638	375.438		
Peri-urban dummy (sq rt)	-3,368.851	2,005.080	-3,959.210	2,113.493		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-597.156**	226.946	-423.257*	206.070		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	-5.545	44.588	-12.820	40.026		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	-110.057	105.235	-83.333	90.484		
People per square km, 2005 (ln, HC) (sq rt)	58.136	661.444	-56.350	604.632		
Number hh members 16-65 (sq rt)	127.851	271.379	-123.128	283.162		
Age of head (sq rt)	-318.914**	112.307	-255.462*	103.124		
Female head (sq rt)	-160.993	237.503	-149.766	227.604		
Yrs educ, adults (ave) (sq rt)	330.882	170.745	306.366*	155.617		
Cost/hired worker (med smth, USD) (sq rt)	21.660	11.099	22.008*	10.709		
Nighttime light ave coverage(DMSP F16) (sq rt)	5.778	17.983	7.653	18.681		
Financial service available (sq rt)	10.006	196.533	12.162	197.103		
Productive non-ag asset ind, fact 1 (sq rt)	2,778.897**	318.084	2,391.215**	326.266		
Productive ag-related asset ind, fact 1 (sq rt)	343.152	3,753.491	1,217.315	3,683.152		
Transfers recvd (USD)					-0.000	0.000
Household size					0.086**	0.008
HH dependent share					-0.377**	0.093
Head's father attended school					0.071	0.048
Years educ, head					0.012*	0.005
Urban					0.529**	0.051
Peri-urban dummy					0.616**	0.101
Ave length of season (days, MOD12Q2)					-0.002	0.001
N	1,985		3,599		3,599	
R2 (adj)	0.219		0.218			
Lambda			-928.40			
Sigma			3786.30			
P value comparison test			0.00			
N (non-censored)			1985			

* $p < 0.05$; ** $p < 0.01$

Table B.7: Comparison of wage employment profit coefficients with and without selection, Generalized Leontief specification. The first model includes only wage market participants (no selection). The second and third columns present results from the second and first stages of a Heckman selection model, respectively. The marginal effects of profit function variables are shown.

	No Select Margins at means	(SE)	Heckman Margins at means	(SE)	Heckman Selection coeffs	(SE)
Urban (sq rt)	-240.086	843.255	-323.012	746.940		
Peri-urban dummy (sq rt)	2,187.252	3,290.466	1,515.902	3,555.486		
Hrs travel to nrst town >500k (LSMS-ISA) (sq rt)	-1.983	383.008	92.147	328.373		
Network dist to nrst town >100k (km, LSMS-ISA) (sq rt)	-45.371	72.980	-64.368	59.958		
Dist nrst major rd (km, LSMS-ISA) (sq rt)	-354.843*	158.223	-237.193*	114.297		
People per square km, 2005 (ln, HC) (sq rt)	1,355.943	1,129.243	348.260	897.252		
Number hh members 16-65 (sq rt)	1,045.595**	341.813	508.537	395.601		
Age of head (sq rt)	58.691	148.335	42.620	145.741		
Female head (sq rt)	-922.449**	308.528	-797.558*	341.084		
Yrs educ, adults (ave) (sq rt)	974.635	553.673	-154.296	624.761		
Max educ in hh (yrs) (sq rt)	1,883.221**	555.952	1,980.455**	586.246		
Nighttime light intensity(DMSP F16) (sq rt)	-46.184	185.853	49.682	163.018		
Returns/ag worker (med smth, USD) (sq rt)	5.589	60.635	-16.314	35.770		
Returns/ind worker (med smth, USD) (sq rt)	3.539	15.342	-2.169	17.224		
Returns/ser worker (med smth, USD) (sq rt)	26.232*	10.957	12.166	11.823		
Partpn in ag emplmt (med smth sh) (sq rt)	231.177	1,331.533	-313.959	1,021.808		
Partpn in ind emplmt (med smth sh) (sq rt)	203.108	709.598	-359.235	1,006.978		
Partpn in ser emplmt (med smth sh) (sq rt)	-491.600	961.593	254.652	803.682		
Transfers recvd (USD)					0.000	0.000
Household size					0.053**	0.008
HH dependent share					-0.797**	0.101
Head's father attended school					0.112*	0.048
Years educ, head					0.049**	0.006
Urban					0.558**	0.051
Peri-urban dummy					0.367**	0.098
Ave length of season (days, MOD12Q2)					0.004**	0.001
N	1,342		3,599		3,599	
R2 (adj)	0.371		0.362			
Lambda			-1697.31			
Sigma			3804.23			
P value comparison test			0.00			
N (non-censored)			1342			

* $p < 0.05$; ** $p < 0.01$

Table B.8: All coefficients, self employment profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_farm_s																				
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	rypot3_low	rypot3_high	tech_use_maize	cpd_farmhire	tractor_share	landowned	rainfall_mean	slope	nutrientretention	soilworkability	irrigation_sh
Level_effect	705	1940	-667	275	406	-1484	1214	479	914	140	11977	-8801	-453	691	660	1234	290	-396	55	702	9870
sd	1217	3701	719	101	187	1185	693	335	709	303	4643	5174	1825	954	2424	524	135	445	1319	1383	7966
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	-9.01	370	230	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	126	445	47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	-31.9	-73.9	-20.3	-1.44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	20.2	51.9	11	1.39	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
dist_road	59.9	-94.2	-30.1	-1.97	-7.57	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	32.2	85.7	18.7	3.2	3.99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
popdensity_ln	-9.75	-113	239	-22.6	-38	-69.6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	322	491	139	23.6	36.5	118	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
pop_indep	-101	257	125	-2.66	-33.9	-163	-31.3	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	121	196	64.4	10.2	15.8	138	45.9	.	.	.	.	.	.	.	.	.	.	.	.	.	.
age_head	-11.7	-27.3	-9.05	2.14	-.988	48.1	-60.7	-14.3	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	59	94.4	30.4	4.65	7.62	65.3	43.1	14.9	.	.	.	.	.	.	.	.	.	.	.	.	.
female_head	-67.6	-101	40.2	-29.3	8.47	126	-96.6	-35.1	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	133	232	75.3	11.8	19.1	161	85.1	43.1	.	.	.	.	.	.	.	.	.	.	.	.	.
educ_yrs_ave	15.5	65.8	42.8	-11.7	2.47	-34.6	4.18	.628	-16	14.3	.	.	.	.	.	.	.	.	.	.	.
sd	72.2	97.1	35	5.44	8.22	67.5	40	17.7	41.9	19	.	.	.	.	.	.	.	.	.	.	.
rypot3_low	1232	1197	468	-166	-79.3	-3762	-460	-21.3	39.3	-38.6	1225	.	.	.	.	.	.	.	.	.	.
sd	1056	1683	550	92.7	158	1198	427	222	530	235	2073	.	.	.	.	.	.	.	.	.	.
rypot3_high	-8.5	-933	-407	109	88.8	3904	314	-249	-301	-187	-1172	122	.	.	.	.	.	.	.	.	.
sd	994	1872	557	90.5	158	1155	453	227	535	245	3568	2493	.	.	.	.	.	.	.	.	.
tech_use_maize	-349	-384	401	19.8	-106	-101	-143	117	-188	96.8	-1526	1486	397	.	.	.	.	.	.	.	.
sd	283	510	165	26.2	50.3	407	170	79.9	197	96.1	1312	1313	494	.	.	.	.	.	.	.	.
cpd_farmhire	-212	29.6	-53	14.1	-42.6	163	61.4	61.5	165	-18.4	-253	92.3	-132	-76.3	.	.	.	.	.	.	.
sd	158	397	87	17.1	25.1	212	89.2	41.2	103	47.7	728	695	214	96.5	.	.	.	.	.	.	.
tractor_share	-153	-1739	-130	71.2	-112	-1018	263	188	-476	234	2217	-172	-708	-406	-1263	.	.	.	.	.	.
sd	543	1167	303	39.3	67.4	659	214	109	265	116	2008	2067	592	456	777	.	.	.	.	.	.
landowned	-214	-237	-55.9	.624	18	-7.11	-85.2	11	127	42.3	230	-12.8	51.2	-221	-131	-26.9	.	.	.	.	.
sd	99.9	148	49.6	8.02	12.4	110	56.4	28.1	71.7	29.5	378	386	128	71.9	154	26	.	.	.	.	.
rainfall_mean	-10.9	-61.1	-30.1	-2.99	-1.82	-.47	-7.88	-1.55	-35.3	-1.15	-104	43.8	7.75	-18.6	-101	-17.7	-1.69	.	.	.	.
sd	24.4	89.7	17.3	2.02	4.14	27.5	13.1	6	14.2	6.81	99.9	103	37.9	18.9	62	11.5	2.07	.	.	.	.
slope	-78.1	-183	30.1	6.87	-12.6	274	-23.9	-18.8	12.6	43.9	271	-71.2	33.7	-25	176	17.2	-.379	-19.1	.	.	.
sd	95.5	234	49.2	7.44	12.4	109	50.2	23.6	57.1	27.1	387	382	135	67.6	201	40	10	22.1	.	.	.
nutrientretention	-204	-572	17.3	-8.49	-66.5	155	-58.5	7.85	342	131	-1261	510	-397	-608	1203	-99.9	-11.2	-254	1176	.	.
sd	235	505	115	20.3	35	237	124	58.5	141	61.9	997	1055	325	201	523	98.4	25	92.1	471	.	.
soilworkability	194	480	-299	-17.9	71.5	141	240	-102	-252	-149	655	-274	-643	212	545	-23.3	59.6	85.3	-968	-192	.
sd	220	570	117	18.8	35.7	247	124	56.6	140	63.1	1030	951	302	182	455	94.9	26.2	84.6	415	261	.
irrigation_sh	-165	10672	584	-213	153	393	84.4	-241	-1677	50.8	-4493	5915	2809	252	-5130	352	-295	-242	1841	-2970	-6172
sd	872	7714	816	146	246	1779	581	280	624	286	4898	4250	1515	825	3068	430	172	444	2118	1456	2788

Table B.9: All coefficients, self employment profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_ent_s														
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	cpw_enthire	lightsum	fs_available	assets_na	assets_ag
Level_effect	-5878	92.8	-3122	-46.6	-231	-4724	-59.6	1951	2466	492	162	65.8	-4551	7854	-31586
sd	3926	5828	2028	387	638	3814	2954	1517	2850	1291	149	134	2501	3717	21403
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	-178	-3121	320	.	.	.	.	.	.	.	.	.	.	.	.
sd	403	1762	180	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	23.5	182	-23.7	8.37	.	.	.	.	.	.	.	.	.	.	.
sd	70.1	244	45.5	6.52	.	.	.	.	.	.	.	.	.	.	.
dist_road	-59.5	197	-31.9	-3.66	-.76	.	.	.	.	.	.	.	.	.	.
sd	136	348	75.7	16.2	19.8	.	.	.	.	.	.	.	.	.	.
popdensity_ln	892	1601	858	28.8	9.23	-789	.	.	.	.	.	.	.	.	.
sd	1433	2198	639	119	204	677	.	.	.	.	.	.	.	.	.
pop_indep	1199	1273	133	15.3	23.2	546	431	.	.	.	.	.	.	.	.
sd	583	979	310	59.7	103	1011	248	.	.	.	.	.	.	.	.
age_head	238	-343	-131	-13.9	59.1	243	-99.5	-106	.	.	.	.	.	.	.
sd	263	476	135	25.4	48.5	483	223	79.2	.	.	.	.	.	.	.
female_head	1044	-740	-202	13.2	-1.13	474	537	-129	.	.	.	.	.	.	.
sd	594	1024	313	62.3	110	960	504	227	.	.	.	.	.	.	.
educ_yrs_ave	-185	-343	20.8	-19.8	34.6	250	-341	-70	-174	-55	.	.	.	.	.
sd	337	505	173	30.5	55.2	461	259	108	258	107	.	.	.	.	.
cpw_enthire	8.47	16.5	16.1	-2.45	-.832	-17	-44.1	-8.24	-45.7	15.3	-.191	.	.	.	.
sd	22.8	81	13.1	2.63	5.22	58.7	20.5	9.19	19.8	10.7	.501	.	.	.	.
lightsum	-60.1	-71	-50.5	1.45	-8.19	62.3	-1.48	2.58	-33.2	2.72	.0389	-1.47	.	.	.
sd	34.7	47.6	19.6	3.57	7.39	54.1	21.1	9.95	20.4	11.8	1.02	.961	.	.	.
fs_available	337	1296	741	22.8	-85.8	1328	136	-15.9	-1150	-6.52	-19.4	2.4	.	.	.
sd	522	881	280	52	95.9	887	441	194	447	237	18.7	20	.	.	.
assets_na	-774	-718	44.6	24.3	-173	-209	-756	-1303	286	694	43.5	-7.33	-229	2001	.
sd	849	1344	407	78.4	152	1271	686	309	742	376	26.7	29.3	601	600	.
assets_ag	8270	-6346	2691	-86	324	9613	-3866	2884	-8241	-2156	-204	-102	3452	-2715	9741
sd	4397	7654	2094	406	673	6185	2834	1608	4262	1909	136	141	2763	4342	12910

Table B.10: All coefficients, wage employment profit first stage regression, Heckman selection model with Generalized Leontief specification

	fs_all_market_s																	
	urban	periurban	traveltime_500k	networkdist_100k	dist_road	popdensity_ln	pop_indep	age_head	female_head	educ_yrs_ave	educ_yrs_max	lightintensity	rpw_ag	rpw_ind	rpw_ser	partshare_ag	partshare_ind	partshare_ser
Level_effect	2515	4395	2256	-707	696	-2777	-4290	-267	-1416	-5579	1461	-1668	455	-35.6	-83.1	-4696	19738	-20070
sd	6473	10077	3390	646	1058	8490	4242	2221	3994	5619	5475	1784	262	203	140	10277	9632	8508
urban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
periurban	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
traveltime_500k	611	2259	-481	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	1024	2806	253	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
networkdist_100k	-97.4	-464	85.5	4.48	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	152	437	66.6	9.54	.	.	.	.	.	.	.	.	.	.	.	.	.	.
dist_road	350	-445	29.7	9.13	6.12	.	.	.	.	.	.	.	.	.	.	.	.	.
sd	256	583	108	23.3	31.7	.	.	.	.	.	.	.	.	.	.	.	.	.
popdensity_ln	-1185	-792	-652	68.5	-45.4	-180	.	.	.	.	.	.	.	.	.	.	.	.
sd	2319	3527	1166	215	331	1153	.	.	.	.	.	.	.	.	.	.	.	.
pop_indep	930	3505	167	1.86	-226	1033	81.8	.	.	.	.	.	.	.	.	.	.	.
sd	889	1427	411	84.1	151	1417	396	.	.	.	.	.	.	.	.	.	.	.
age_head	-182	191	214	-28.3	11.7	579	62.2	-136	.	.	.	.	.	.	.	.	.	.
sd	385	622	182	36.6	71	671	302	105	.	.	.	.	.	.	.	.	.	.
female_head	-231	-802	88.8	32.1	-90.9	1202	381	-274	.	.	.	.	.	.	.	.	.	.
sd	842	1472	388	75.4	154	1321	655	303	.	.	.	.	.	.	.	.	.	.
educ_yrs_ave	1676	3762	-28	-130	-18.4	.939	-165	-22.2	1228	680	.	.	.	.	.	.	.	.
sd	1159	1608	513	106	192	1939	947	447	963	968	.	.	.	.	.	.	.	.
educ_yrs_max	-813	-2799	-46.2	96.5	150	86.2	697	78	-1144	1155	-680	.	.	.	.	.	.	.
sd	1152	1660	512	100	188	1900	919	419	933	2007	1136	.	.	.	.	.	.	.
lightintensity	64.4	-422	-104	43.8	2.1	1444	-231	-175	172	-670	610	-129	.	.	.	.	.	.
sd	390	766	253	38.8	67.9	700	282	129	283	388	379	110	.	.	.	.	.	.
rpw_ag	111	-157	-60.2	.333	-3	-106	-47.7	-14.9	17.6	-4.96	6.11	-5.27	2.02	.	.	.	.	.
sd	74	136	57.4	6.38	10.9	83.2	33.5	15.9	35.8	43.4	43.1	16.2	1.97	.	.	.	.	.
rpw_ind	-60.8	-69.4	-10.8	6.56	-21.5	-36.9	31.4	5.31	2.13	-32.5	13.8	-5.47	-2.76	3.04	.	.	.	.
sd	43.6	57.6	17.5	4.78	7.65	69.7	25.6	11.9	25.3	32	32.4	13.2	1.75	1	.	.	.	.
rpw_ser	-40.9	-37.6	14.8	1.9	-8.35	-31.4	50.1	3.52	-4.3	38.2	-35.6	11.5	1.83	-1.3	.335	.	.	.
sd	31.1	65.5	16	3.11	5.26	48.5	15.6	6.76	15.2	22.1	23.2	9.03	1.28	1.02	.251	.	.	.
partshare_ag	-3012	-9336	-897	151	36.1	917	-1272	425	835	758	-672	418	47.4	10.8	-1.72	-1578	.	.
sd	2291	3701	1232	191	335	3504	1565	709	1631	1979	1948	732	82.9	81.6	46.6	3486	.	.
partshare_ind	1660	-1431	-434	-119	-126	-6523	209	-732	-2733	-1807	730	-407	-97.2	164	-7.36	-4517	5943	.
sd	2106	3256	907	172	383	3476	1490	673	1396	1908	1879	701	92.3	68.7	40.9	3545	4060	.
partshare_ser	-4153	3278	376	136	63.3	5179	-3707	1688	-2848	2456	-3060	-47.4	45.9	-12.7	66.8	6729	1218	300
sd	2280	3264	902	188	330	3159	1558	720	1547	1962	1882	697	94.1	64.8	45.7	3597	3438	2470

BIBLIOGRAPHY

Anderson, J.R., and J.B. Hardaker. 2003. "Risk aversion in economic decision making: Pragmatic guides for consistent choice by natural resource managers." In J. Wesseler, H. P. Weikard, and R. D. Weaver, eds. *Risk and Uncertainty in Environmental Economics*. Gloucestershire, UK: Edward Elgar Publishing, pp. 171–188.

Arthi, V., K. Beegle, J.D. Weerdt, and A. Palacios-Lopez. 2016. "Measuring Household Labor on Tanzanian Farms." Unpublished.

Banerjee, A.V., and A.F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101:274–298.

Bänziger, M., P.S. Setimela, D. Hodson, and B. Vivek. 2006. "Breeding for improved abiotic stress tolerance in maize adapted to southern Africa." *Agricultural Water Management* 80:212–224.

Barrett, C. 2005. "Rural poverty dynamics: development policy implications." *Agricultural Economics*, pp. .

Barrett, C.B. 2007. "Displaced distortions: Financial market failures and seemingly inefficient resource allocation in low-income rural communities." In E. Bulte and R. Ruben, eds. *Development Economics Between Markets and Institutions: Incentives for growth, food security and sustainable use of the environment.*. Mansholt publication series, Volume 4. Wageningen: Wageningen Academic Publishers.

Barrett, C.B., T. Reardon, and P. Webb. 2001. "Nonfarm income diversification and household livelihood strategies in rural Africa: concepts, dynamics, and policy implications." *Food Policy* 26:315–331.

Beegle, K., C. Carletto, and K. Himelein. 2012. "Reliability of recall in agricultural data." *Journal of Development Economics* 98:34–41.

Beegle, K., J.D. Weerdt, and S. Dercon. 2011. "Migration and economic mobility in Tanzania: Evidence from a tracking survey." *Review of Economics and Statistics* 93:1010–1033.

Behrman, J. 1999. "Labor markets in developing countries." *Handbook of Labor Econonomics, volume 3* 3:2859–2939.

Benin, S., and B. Yu. 2013. "Complying with the Maputo Declaration Target: Trends in Public Agricultural Expenditures and Implications for Pursuit of Optimal Allocation of Public Agricultural Spending." Working paper, IFPRI.

Bewket, W. 2009. "Rainfall variability and crop production in Ethiopia: Case study in the Amhara region." In *Proceedings of the 16th International Conference of Ethiopian Studies*. Norwegian University of Science and Technology Trondheim, Norway, pp. 823–836.

Bhalla, S.S. 1978. "The role of sources of income and investment opportunities in rural savings." *Journal of Development Economics* 5:259–281.

Binswanger-Mkhize, H.P., and S. Savastano. 2014. "Agricultural intensification: the status in six African countries." Unpublished.

Block, S.A. 2013. "The post-independence decline and rise of crop productivity in sub-Saharan Africa: measurement and explanations." *Oxford Economic Papers*, mar, pp. .

Bomba, K. 2016. "Learn as you go: the Ethiopian example of agricultural transformation in action." *Foreign Affairs*, pp. .

Boserup, E. 1990. *Economic and demographic relationships in development*. The Johns Hopkins studies in development, Baltimore, MD: Johns Hopkins University Press.

Bound, J., C. Brown, and N. Mathiowetz. 2001. "Measurement Error in Survey Data." In J. J. Heckman and B. T. Leamer, eds. *Handbook of Econometrics*. Elsevier, vol. Volume 5, pp. 3705–3843.

Burke, M.B., D.B. Lobell, and L. Guarino. 2009. "Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation." *Global Environmental Change* 19:317–325.

Bustos, P., B. Caprettini, and J. Ponticelli. 2012. "Agricultural productivity and structural transformation. Evidence from Brazil." Unpublished.

Chamberlin, J., and E. Schmidt. 2012. "Ethiopian agriculture: A dynamic geographic perspective." In P. Dorosh and S. Rashid, eds. *Food and Agriculture in Ethiopia: Progress and Policy Challenges*. Philadelphia: University of Pennsylvania Press, pp. 21–52.

Chavas, J.P., and M.T. Holt. 1996. "Economic Behavior Under Uncertainty: A Joint Analysis of Risk Preferences and Technology." *Review of Economics and Statistics* 78:329–335.

Christiaensen, L., L. Demery, and J. Kuhl. 2011. "The (evolving) role of agriculture in poverty reduction - An empirical perspective." *Journal of Development Economics* 96:239–254.

CIMMYT Economics Program. 1988. *From agronomic data to farmer recommendations: an economics training manual*. 27, CIMMYT (free PDF download).

Collier, P., and S. Dercon. 2014. "African agriculture in 50 years: smallholders in a rapidly changing world?" *World Development* 63:92–101.

Davis, B., S. Di Giuseppe, and A. Zezza. 2014. "Income Diversification Patterns in Rural Sub-Saharan Africa Reassessing the Evidence." *World Bank Policy Research Working Paper* No. 7108.

Davis, B., P. Winters, G. Carletto, K. Covarrubias, E.J. Quiñones, A. Zezza, K. Stamoulis, C. Azzarri, and S. DiGiuseppe. 2010. "A Cross-Country Comparison of Rural Income Generating Activities." *World Development* 38:48–63.

De Janvry, A., M. Fafchamps, and E. Sadoulet. 1991. "Peasant household behaviour with missing markets: some paradoxes explained." *Economic Journal* 101:1400–1417.

De Janvry, A., and E. Sadoulet. 2010. "Agricultural Growth and Poverty Reduction : Additional Evidence." *World Bank Reserach Observer* 25:1–20.

de Mel, S., D.J. McKenzie, and C. Woodruff. 2009. "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Economics* 88:19–31.

Dercon, S. 2013. "Agriculture and development: revisiting the policy narratives." *Agricultural Economics* 44:183–187.

—. 2009. "Rural Poverty: Old Challenges in New Contexts." *World Bank Research Observer* 24:1–28.

Dercon, S., and L. Christiaensen. 2011. "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia." *Journal of Development Economics* 96:159–173.

Dercon, S., and D. Gollin. 2014. "Agriculture in African Development: Theories and Strategies." *Annual Review of Resource Economics* 6:471–492.

Dillon, B., and C.B. Barrett. 2014. "Agricultural Factor Markets in Sub-Saharan Africa An Updated View with Formal Tests for Market Failure." *World Bank Policy Research Working Paper* No. 7117.

Eswaran, M., and A. Kotwal. 1985. "A theory of two-tier labor markets in agrarian economies." *American Economic Review*, pp. 162–177.

Ethiopian Ministy of Finance and Economic Development. 2010. "Ethiopia Growth and Transformation Plan (GTP)."

Evenson, R.E., and D. Gollin. 2003. "Assessing the Impact of the Green Revolution, 1960 to 2000." *Science* 300:758–762.

Feder, G., R.E. Just, and D. Zilberman. 1985. "Adoption of agricultural innovations in developing countries: A survey." *Economic development and cultural change* 33:255–298.

Foster, A., and M. Rosenzweig. 2003. "Agricultural development, industrialization and rural inequality." Unpublished.

—. 2007. "Economic development and the decline of agricultural employment." In T. P. Schultz and J. Strauss, eds. *Handbook of Development Economics*. Amsterdam: North-Holland, vol. 4, chap. 47, pp. 3051–3083.

Gollin, D. 2014. "The Lewis Model: A 60-Year Retrospective." *The Journal of Economic Perspectives* 28:71–88.

Gollin, D., D. Lagakos, and M.E. Waugh. 2014a. "Agricultural Productivity Differences across Countries." *American Economic Review* 104:165–170.

—. 2014b. “The Agricultural Productivity Gap.” *Quarterly Journal of Economics* 129:939–993.

Gollin, D., S. Parente, and R. Rogerson. 2002. “The Role of Agriculture in Development.” *The American Economic Review* 92:160–164.

Gu, Y., A.R. Hole, and S. Knox. 2013. “Fitting the generalized multinomial logit model in Stata.” *Stata Journal* 13:382–397.

Haefele, S.M., K. Naklang, D. Harnpichitvitaya, S. Jearakongman, E. Skulkhu, P. Romyen, S. Phasopa, S. Tabtim, D. Suriya-arunroj, S. Khunthasuvon, D. Kraisorakul, P. Youngsuk, S.T. Amarante, and L.J. Wade. 2006. “Factors affecting rice yield and fertilizer response in rainfed lowlands of northeast Thailand.” *Field Crops Research* 98:39–51.

Haggblade, S., P. Hazell, and T. Reardon. 2010. “The Rural Non-farm Economy: Prospects for Growth and Poverty Reduction.” *World Development* 38:1429–1441.

Haggblade, S., P.B. Hazell, and T.A. Reardon. 2007. “Structural Transformation of the Rural Nonfarm Economy.” In S. Haggblade, P. B. Hazell, and T. A. Reardon, eds. *Transforming the rural nonfarm economy: opportunities and threats in the developing world*. Baltimore, Md.: Johns Hopkins University Press, chap. 4, p. Chapter 4.

Harou, A., Y. Liu, C.B. Barrett, and L. You. 2013. “Poverty rates and the returns to fertilizer: empirical and simulation evidence from Malawi.”, pp. .

Harris, J.R., and M.P. Todaro. 1970. “Migration, unemployment and development: a two-sector analysis.” *The American Economic Review* 60:126–142.

Hayami, Y., and V.W. Ruttan. 1971. *Agricultural development; an international perspective*. Baltimore: Johns Hopkins Press.

Headey, D., and T. Jayne. 2014. "Adaptation to land constraints: Is Africa different?" *Food Policy* 48:18–33.

Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47:153–161.

Heisey, P.W., and W. Mwangi. 1997. "Fertilizer use and maize production." *Africa's Emerging Maize Revolution*, pp. 193–212.

Irz, X., and T. Roe. 2005. "Seeds of growth? Agricultural productivity and the transitional dynamics of the Ramsey model." *European Review of Agricultural Economics* 32:143–165.

Jayne, T.S., and S. Rashid. 2013. "Input subsidy programs in sub-Saharan Africa: a synthesis of recent evidence." *Agricultural Economics* 44:547–562.

Johnston, B., and J. Mellor. 1961. "The role of agriculture in economic development." *American Economic Review* 51:566–593.

Jolliffe, D. 2004. "The impact of education in rural Ghana: examining household labor allocation and returns on and off the farm." *Journal of Development Economics* 73:287–314.

Keane, M., and R. Moffitt. 1998. "A structural model of multiple welfare program participation and labor supply." *International Economic Review* 39:553–589.

Keane, M.P., and K.I. Wolpin. 1997. "The Career Decisions of Young Men." *Journal of Political Economy* 105:473–522.

Korecha, D., and A.G. Barnston. 2007. "predictability of June-September rainfall in Ethiopia." *Monthly weather review* 135:628–650.

Lagakos, D., and M. Waugh. 2013. "Selection, Agriculture and Cross-Country Productivity Differences." *American Economic Review* 103:948–980.

Laver, M., K. Benoit, and J. Garry. 2003. "Extracting policy positions from political texts using words as data." *American Political Science Review* 97:311–331.

Lele, U., M. Agarwal, and S. Goswami. 2013. "Lessons of the Global Structural Transformation Experience for the East African Community." Unpublished.

Lewis, W.A. 1954a. "A model of dualistic economics." *American Economic Review* 36:46–51.

—. 1969. *Aspects of tropical trade 1883-1965*. Almqvist & Wiksell (distr.).

—. 1954b. "Unlimited Supplies of Labour." *Manchester school*, pp. .

Lobell, D.B., M. Banziger, C. Magorokosho, and B. Vivek. 2011. "Nonlinear heat effects on African maize as evidenced by historical yield trials." *Nature Clim. Change* 1:42–45.

Magdoff, F., and H. Van Es. 2000. *Building Soils for Better Crops.*, 2nd ed. Beltsville, MD: Sustainable Agriculture Network.

Marenya, P.P., and C.B. Barrett. 2009. "State-conditional fertilizer yield response on western Kenyan farms." *American Journal of Agricultural Economics* 91:991–1006.

McCaig, B., and N. Pavcnik. 2013. "Moving out of Agriculture: Structural Change in Vietnam." *National Bureau of Economic Research Working Paper Series* No. 19616.

McCullough, E. 2015. "Labor productivity and employment gaps in Sub-Saharan Africa." *Policy Research Working Paper No. 7234*, pp. .

McMillan, M., and K. Harttgen. 2014. "What is driving the African Growth Miracle'?" *National Bureau of Economic Research Working Paper Series* No. 20077.

McMillan, M., and D. Headey. 2014. "Introduction Understanding Structural Transformation in Africa." *World Development* 63:1–10.

McMillan, M., and D. Rodrik. 2011. "Globalization, Structural Change and Productivity Growth." *National Bureau of Economic Research Working Paper Series* No. 17143.

McMillan, M., D. Rodrik, and I. Verduzco. 2014. "Globalization, Structural Change and Productivity Growth, with an Update on Africa." *World Development* 63:11–32.

Michaels, G., F. Rauch, and S.J. Redding. 2012. "Urbanization and Structural Transformation." *Quarterly Journal of Economics* 127.

Morris, M., V.A. Kelly, R.J. Kopicki, and D. Byerlee. 2007. *Fertilizer use in African agriculture: Lessons learned and good practice guidelines*. Washington, DC: World Bank.

Nagler, P., and W. Naudé. 2014. "Non-Farm Enterprises in Rural Africa: New Empirical Evidence." *World Bank Policy Research Working Paper* No. 7066.

Nelson, L.A., R.D. Voss, and J. Pesek. 1985. "Agronomic and statistical evaluation of fertilizer response." In O. Engelstad, ed. *Fertilizer Technology and Use*. Madison, WI: Soil Science Society of America, pp. 53–90.

Osgood, D., M. McLaurin, M. Carriquiry, A. Mishra, F. Fiondella, J. Hansen, N. Peterson, and N. Ward. 2007. "Designing Weather Insurance Contracts for Farmers in Malawi, Tanzania and Kenya." *Final Report to The Commodity Risk Management Group, ARD, World Bank*, pp. .

Palacios-Lopez, A., L. Christiaensen, and T. Kilic. 2015. "How much of the labor in African agriculture is provided by women?" Unpublished.

Pardey, P.G. 2014. "African Agricultural R&D and Productivity Growth in a Global Setting." In W. P. Falcon and R. Naylor, eds. *Frontiers in Food Policy: Perspectives on Sub-Saharan Africa*. Stanford Center on Food Security and the Environment.

Ranis, G. 2004. "The Evolution of Development Thinking: Theory and Policy." *Yale University Economic Growth Center Discussion Paper 886*, pp. 40pp.

Rashid, S., N. Tefera, G. Ayele, and G.T. Abate. 2012. "Fertilizer in Ethiopia: Policies, Value Chain, and Profitability." Working paper, IFPRI.

Reardon, T., J. Berdegué, C.B. Barrett, and K. Stamoulis. 2006. "Household Income Diversification into Rural Nonfarm Activities." In S. Haggblade, P. Hazell, and T. Reardon, eds. *Transforming the Rural Nonfarm Economy*. Baltimore, MD: Johns Hopkins University Press.

Rodrik, D. 2014a. "An African Growth Miracle?" *National Bureau of Economic Research Working Paper Series* No. 20188.

—. 2014b. "An African Growth Miracle." Unpublished.

Ruthenberg, H. 1971. *Farming systems in the tropics.*. Oxford: Clarendon Press.

Sarkar, A.N., and R.G. Wynjones. 1982. "Effect of rhizosphere pH on the availability and uptake of Fe, Mn and Zn." *Plant and Soil* 66:361–372.

Schlede, H. 1989. "Distribution of acid soils and liming materials in Ethiopia. Ethiopian Institute of Geological Survey." Working paper.

Schultz, T.W. 1964. *Transforming traditional agriculture*. New Haven, CT, USA: Yale Univ. Press.

Sheahan, M., and C.B. Barrett. 2014. "Understanding the agricultural input landscape in sub-Saharan Africa: Recent plot, household, and community-level evidence." *World Bank Policy Research Working Paper*, pp. .

Sherlund, S.M., C.B. Barrett, and A.a. Adesina. 2002. "Smallholder technical efficiency controlling for environmental production conditions." *Journal of Development Economics* 69:85–101.

Snapp, S., T.S. Jayne, W. Mhango, T. Benson, and J. Ricker-Gilbert. 2014. "Maize-Nitrogen Response in Malawi's Smallholder Production Systems." In *National Symposium on Eight Years of FISPImpact and What Next*. Citeseer, pp. 14–15.

Spielman, D.J., D. Kelemwork, and D. Alemu. 2011. "Seed, fertilizer, and agricultural extension in Ethiopia.", pp. .

Tadesse, M., B. Alemu, G. Bekele, T. Tebikew, J. Chamberlin, and T. Benson. 2006. "Atlas of the Ethiopian rural economy." *Ethiopian Development Research Institute*, pp. .

Timmer, C.P. 2009. *A world without agriculture: the structural transformation in historical perspective*. Washington, D.C.: AEI Press.

—. 1988. "The agricultural transformation." In H. B. Chenery and T. N. Srinivasan, eds. *Handbook of development economics, Volume 1*. Amsterdam: Elsevier, pp. 275–331.

Train, K.E. 2002. *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge: Cambridge University Press.

Uyovbisere, E.O., and G. Lombim. 1991. "Efficient fertilizer use for increased crop production: The sub-humid Nigeria experience." *Fertilizer research* 29:81–94.

van Soest, A., M. Das, and X. Gong. 2002. "A structural labour supply model with flexible preferences." *Journal of Econometrics* 107:345–374.

Vanlauwe, B., and K. Giller. 2006. "Popular myths around soil fertility management in sub-Saharan Africa." *Agriculture, Ecosystems & Environment* 116:34–46.

Vollrath, D. 2014. "The efficiency of human capital allocations in developing countries." *Journal of Development Economics* 108:106–118.

Walker, W.E., P. Harremoës, J. Rotmans, J.P. van der Sluijs, M.B.A. van Asselt, P. Janssen, and M.P. Krayer von Krauss. 2003. "Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support." *Integrated Assessment* 4:5–17.

World Bank. 2008. *World Development Report 2008: Agriculture for Development*. Washington, D.C.: World Bank.

World Bank Group. 2014. "World Development Indicators."

Yanggen, D., V. Kelly, T. Reardon, and A. Naseem. 1998. "Incentives for fertilizer use in sub-Saharan Africa: A review of empirical evidence on fertilizer response and profitability." Unpublished.